We thank the reviewers for their careful reading of our paper; their comments have led to several improvements and corrections. In this revision, we have addressed all the comments raised by the reviewers, and we provide a point-to-point response to each comment of the review team. Reviewer comments are in black, our responses are in green.

Reviewer 2

The paper reformulates forecast reconciliation as a grouped variable selection problem. Regularisation and subset selection are carried out using state-of-the-art optimisation techniques, including Mixed Integer Programming. Recognising that reconciled forecasts $\tilde{y}$ are simply linear combinations of base forecasts $\hat{y}$ via $\tilde{y} = SG\hat{y}$, setting columns of G to zero can eliminate heavily misspecified forecasts from the combination. The simulation results and empirical examples are thorough. In so far as selection methods work, they do so when a diagonal matrix is plugged in to the objective function in place of the forecast covariance matrix.

Main Comments

- The discussion around Proposition 1 is confusing. The discussion mixes up the roles of 1) the constraint $GS = I$ and 2) the property of preserving the unbiasedness of forecasts after reconciliation. The proposition as it is currently phrased states: *"Under the assumption of unbiasedness, the count of nonzero column entries of G,... derived through solving equation 4 is at least equal to the number of time series at the bottom level."*. However, equation 4 is solved with the constraint $GS = I$. This is what guarantees that unbiased base forecasts will remain unbiased after reconciliation. The wording implies that this reasoning operates the other way around.

  A further implication of imposing the constraint $GS = I$ is that G must have no less than n_b non-zero columns, as is correctly argued in the proof to the proposition. However, an alternative and more precise wording of Proposition 1 would be *"If the assumption that forecast reconciliation preserves unbiasedness is imposed by enforcing $GS = I$, then the number of nonzero column entries of $\hat{G}$ will be no less than n_b"*.

  Thank you for these helpful comments. We have refined the wording of Proposition 1 as per your suggestions.

- The second part of the Proposition 1 states *"In addition, we can restore the full hierarchal structure by aggregating/disaggregating the selected time series"*. This again is somewhat imprecise and possibly refers to a different issue to that proven here. I suspect that what is meant by 'restore' here, is the following. If the solution to Equation 4 yields a $\hat{G}$ with exactly n_b non-zero columns, then these correspond to variables from which the full hierarchy can be obtained using nothing but the information embedded in the constraints. As the authors suggest, it may be possible that the zero columns correspond to series AA and BA, but not to series AA and AB, since in the later case, the aggregation constraints alone are insufficient for forecasts of AA and AB to be recovered. This is not a consequence of "assuming unbiasedness" or even that $\hat{G}_{.\mathbb{S}}$ has n_b columns. It is a consequence of enforcing the constraint $GS = I$.

  To this I would add the following observations. First, solving equation (4) could lead to a solution for $\hat{G}$ that has more than n_b non-zero columns. In this case there are in fact too many series and a while a coherent forecast can be 'restored', this cannot be done uniquely. Second, is the point that it is the constraint $GS = I$ that enforces that the selected columns of $\hat{G}$ will correspond to variables that can restore the hierarchy. The mechanism by which it does so, is not rigorously proven here, however, a proof should be possible by leaning on some of the arguments made in Zhang et al. (2023), which is cited by the authors.

  We have now corrected Proposition 1 to make it clearer.

Regarding the proof (located in Appendix A.1 of the revised version), we have added explanations to address the two possible cases: one resulting in a $\hat{\boldsymbol{G}}$ matrix with exactly n_b nonzero columns, and the other resulting in a $\hat{\boldsymbol{G}}$ with more than n_b nonzero columns. Additionally, we have rewritten the proof of the second part of Proposition 1 and applied Theorem 2 from Zhang et al. (2023) to substantiate it.

- As well as the statement of Proposition 1, the proof could be made tighter and clarified. In particular the same equation is essentially presented twice, once with $\hat{\boldsymbol{G}}$ and the second time with $\hat{\boldsymbol{G}}_{.\mathbb{S}}$. Only the second of these is needed.

  We have streamlined the proof of Proposition 1 and eliminated the redundant equations.

- Finally, in terms of tightening up the mathematics on page 8, at the very bottom, the line vec($\hat{\boldsymbol{y}}$) = $\hat{\boldsymbol{y}}$, while correct, adds nothing since vec($\hat{\boldsymbol{y}}$) is not used in equation (5). Also, on the second line, it would be worthwhile to make it explicit that $\boldsymbol{SG\hat{y}}$ = vec($\boldsymbol{SG\hat{y}}$) = ($\hat{\boldsymbol{y}} \otimes \boldsymbol{S}$) vec($\boldsymbol{G}$), currently the second and third terms are present but not the first.

  Regarding the proof of Proposition 2 (located in Appendix A.2 of the revised version), we have removed the line vec($\hat{\boldsymbol{y}}$) = $\hat{\boldsymbol{y}}$ and made the final equation explicit, as suggested.

- Some of the methods discussed in the literature review are 'in-sample' methods in the sense that $\hat{\boldsymbol{y}}_{t+h|t}$ are predictions in the form of fitted values ($\boldsymbol{y}_{t+h}$ is in training data when base forecasts are computed). Others (for example the RERM method) are 'out-of-sample' in the sense that $\hat{\boldsymbol{y}}$ are genuine forecasts. In principle all optimisation methods could use either an in-sample or out-of-sample approach. I believe that in this paper only 'in-sample' methods are considered. This is reasonable, however, this should be clearly stated at some point (and it would provide motivation for not using RERM in the simulation studies and empirical results).

  Thank you for the suggestion. We had initially stated the information used by the proposed methods (whether in-sample or out-of-sample) briefly in the fifth paragraph of Section 1 and the first paragraph of Section 3. We have now expanded on this by incorporating your points into both the literature review (Section 2.2) and the methodology section (Section 3).

  We have also added one paragraph (the third paragraph in Section 3.2) explaining why we excluded the RERM and ERM methods from simulation studies and empirical results. Specifically, these "out-of-sample" methods rely on an iterative approach with expanding windows to generate forecasts, demanding extensive rounds of model training and significant computation time. In contrast, the methods we included in the results require only a single round of training and forecasting. Thus, RERM and ERM were not considered in the results.

- More guidance could be given on the similarities between methods. For example Elasso seems to be the same as OLS-lasso with the important difference that only the latter retains the $\boldsymbol{GS} = \boldsymbol{I}$ constraint. This also begs two further questions. The first is why a $\boldsymbol{W}$ matrix is not used in the Elasso method. The second is why the $\boldsymbol{GS} = \boldsymbol{I}$ constraint is not dropped for other regularisation approaches.

  Thank you for the helpful comments. We have revised the paragraph following Proposition 2 to clarify that the three constrained "out-of-sample" reconciliation methods differ only in their regularization terms used for series selection. We have also added a new paragraph (the second paragraph of Section 3.2) to explain the difference between Elasso and the three constrained reconciliation methods. Specifically, Elasso differs from OLS-lasso in two main aspects: the inclusion of the constraint $\boldsymbol{GS} = \boldsymbol{I}$ and the data sources used (i.e. OLS-lasso is an "out-of-sample" method, while Elasso is an "in-sample" method).

  To address the two questions raised, we have added Equation (2) to show the decomposition of the loss function considered in forecast reconciliation methods. The Elasso method, operating as an "in-sample" method, relaxes the unbiasedness assumption and omits the $\boldsymbol{GS} = \boldsymbol{I}$ constraint,

aiming to minimize the mean squared reconciled forecast errors directly, as shown in Equation (2). However, the other regularization approaches, which are "out-of-sample", do not have access to the true values in the test set. Therefore, they assume unbiasedness and impose the $GS = I$ constraint to cancel the bias term in Equation (2) and focus on minimizing only the variance term in Equation (2). We have made several edits to reflect these clarifications; see, for example, the second paragraph of Section 3.1 and the second paragraph of Section 3.2.

We acknowledge the possibility to include a W matrix in Elasso. However, in this context, the W matrix can be arbitrary and is not restricted to the covariance matrix of base forecast errors, as defined in the paper. The purpose of the W matrix here is to assign different weights to reconciled forecast errors of various series within the a hierarchy, unlike our approaches, which treat reconciled forecast errors equally across all series. Therefore, we have highlight this concept as potential future work and have included it in the newly added discussion section.

- Where the unbiasedness preserving property is dropped, the authors could also consider including an n_b-dimensional shift parameter d, such that $\tilde{y} = S(d + G\tilde{y})$. Then d can be trained alongside with vec(G) and act as a bias correction. The problem should still be able to be written down as a least squares problem, optimising w.r.t. (d, vec(G)).

Thank you for your valuable comment. In the Elasso method, we drop the unbiasedness assumption and directly minimize the loss function as defined in Equation (2). So we did not incorporate a bias correction in order to maintain consistency across the proposed methods. The suggestion of including a shift parameter, which should be of the form $D \in \mathbb{R}^{T \times n_b}$, is insightful. We have added a paragraph in the discussion section to further elaborate on this idea.

- In the intuitive method it is on the one hand stated that '*when the jth diagonal element of A is zero, the jth column of $\bar{G}$ becomes entirely composed of zeros*'. However, later it is stated that *"implementing grouped variable selection… can be challenging because it imposes restrictions of $\bar{G}$ to ensure it adheres rigorously to the analytical solution of MinT while making the selection"*. The second, quite confusing statement, seems to contradict the first, if a zero element in A implies a zero column of $\bar{G}$ then why is grouped selection even necessary? Also, when calling a method 'intuitive' it is important to discuss what makes it intuitive. Little intuition is given to motivate this method, rather an appeal is made to reduce the number of parameters - perhaps 'parsimonious' method would be a more appropriate name.

We have now revised the first two paragraphs related to the Parsimonious method to correct the statements and improve clarity. Also, we have renamed the "Intuitive method" to the "Parsimonious method" throughout the revised manuscript and updated all corresponding results accordingly.

- On page 24 the statement is made that *"the Elasso method consistently outperforms the others overall"*. Given that the Elasso performs poorly at short horizons and for some groups of bottom level variables, I think the use of 'consistently' is not warranted here.

We have now fixed it by completely rewriting the last paragraph of Section 5.1 and have thoroughly reviewed the paper to ensure the accuracy of statements.

- The results for most methods are very close to one another. Some testing on whether the observed differences are significant should be added.

Thanks for the comment. We have now included the MCB test results for the two simulations and have updated the related results analysis, see Section 4 and Appendices A-B of the supplementary material. We note that the MCB tests were not conducted for the two empirical applications due to limited data length and insufficient test sets; we have added explanations for this in the applications section.

- While it is valuable to report the series that tend to be selected more often, some additional context would be useful - in particular the forecast variance of each series (in order to determine whether series with high forecast error are dropped) and the forecast correlation (to determine whether uncorrelated series are selected).

    Thank you for the insightful comment. In hierarchical forecasting, higher-level time series generally have greater forecast error variance compared to lower levels. To address this, we use the scaled error metric, RMSSE, to assess the scaled forecast variance, enabling comparison across series from different levels. We have now reported the RMSSE values of base forecasts for each series in both simulations, specifically in Tables 3, B.2, and B.4 for simulation setup 1, and Tables 5 and C.1-C.4 for simulation setup 2. We have also updated the related results analysis to incorporate insights from the RMSSE findings, where it was generally observed that series with higher RMSSE values are selected less frequently. For the two empirical applications, due to the large number of time series in each hierarchy, reporting RMSSE values for each series individually is impractical. Therefore, we have reported the average RMSSE values at each level to compare whether the average RMSSE for selected series is lower than that of all series at the same level. For most cases, the series selected by the proposed methods show reduced RMSSE values at each level.

    Regarding forecast correlation, we have included heatmaps showing the correlations between base forecast errors for both simulations. These are presented in Figures 2(b), B.1(b), and B.2(b) for setup 1, and Figures 4(b) and C.1(b) for setup 2. We have also expanded the related results analysis to incorporate insights from these results. Our analysis suggests that both forecast variance and forecast correlation influence time series selection. In general, time series with high scaled forecast variance and strong correlations with other series are more likely to be excluded. For the empirical applications, we did not include heatmaps, as the number of time series in these hierarchies is too large to present them effectively.

Reviewer 3

This paper proposes novel forecast reconciliation methods incorporating time series selection. Two categories of such methods are proposed - one category is based on out-of-sample information while the other category of methods is based on in-sample information. These are illustrated via simulation studies and two empirical applications. The findings are argued to demonstrate improved forecast accuracy, especially at higher aggregation levels, longer forecast horizons and in situations with model misspecification.

The paper focuses on an important area (i.e., forecast reconciliation), is theoretically rigorous and provides a good summary of the relevant research. Incorporating time series selection to strengthen forecast reconciliation has both theoretical and practical significance. While the authors are to be commanded on their theoretical approach, the paper lacks any substantive discussion on the practical significance of the proposed methods. What are the precise implications of their proposed methods for analysts working in labour economics and tourism, for example (since their application data is from these two fields). Other users of such forecasts? What could be some potential implications for decision making? Policy making? These would be vital to enhancing the paper's reach and impact.

    Thank you for the comment. We have now added few more lines to the third paragraph of the introduction section, and included an additional paragraph in both the introduction and discussion sections to discuss the potential practical significance of the proposed methods. Below, we only present the newly added content from the introduction, with the key practical significance highlighted in bold.

    "In practice, hierarchical time series can encompass hundreds or even thousands of individual series, making it impractical to focus on each model we fit. To address this complexity, we often rely on automated model selection methods."

"Unlike Zhang et al. (2023), which keep good forecasts unchanged, we leave poor base forecasts unused in generating reconciled forecasts. Notably, **our methods automate the selection process, relieving practitioners from the task of carefully choose which series to exclude.** There may be other practical motivations to use our proposed approach. For example, in labour force forecasting, economists may **be interested in various attributes** such as gender, labor market region, and duration of job search, which create a complex hierarchical structure. **Some series within this structure may be poorly specified for forecasting or exhibit weak patterns**, resulting in poor base forecasts. **Our methods selectively use some series into the reconciliation process, potentially improving overall accuracy while still providing reconciled forecasts for the entire hierarchy.** Moreover, **the selection procedure avoids the need to forecast the excluded series in future operations, at least for a short duration**, and employs only the selected series along with the estimated weighting matrix for reconciliation, thereby improving computational efficiency. Finally, **they also adapt well to scenarios with missing or extremely sparse data** by adding additional constraints to exclude such series, enabling us to bypass their base forecasts while still returning reconciled outcomes."

- P.2 The authors claim "Through simulation experiments and two empirical applications, we demonstrate that our proposed methods guarantee coherent forecasts that outperform or match their respective benchmark methods" - how is such a guarantee provided? May wish to reconsider the wording.

  Thanks for pointing this out. We have now completely rewritten the fifth paragraph in the introduction section and thoroughly reviewed the paper to ensure the accuracy of statements.

- p.2-3 "A remarkable feature of the proposed methods is their ability to diminish disparities arising from using different estimates of the base forecast error covariance matrix, thereby mitigating challenges associated with estimator selection, which is a prominent concern in the field of forecast reconciliation research." What about other concerns in forecast reconciliation research? It would be good to summarize these and address areas where the proposed methodology could support such concerns.

  In the newly added paragraph in the introduction section, we also have discussed other concerns in forecast reconciliation research. For example, in complex hierarchical structures, some series may be poorly specified for forecasting or exhibit weak patterns, leading to suboptimal base forecasts. Additionally, forecasting all series in such complex hierarchies can be time-consuming. Although our methods can be time-consuming in performing reconciliation with time series selection, they offer a possible advantage once implemented. The selection procedure embedded in our methods avoids the need to forecast the excluded series in future operations, at least for a short duration, and employs only the selected series along with the estimated weighting matrix for reconciliation, thereby improving computational efficiency. Lastly, in scenarios with missing or extremely sparse data, it may be necessary to exclude these series from the reconciliation process. Our proposed methods are designed to adapt effectively to these scenarios.

- Conclusion section needs to be extended to discuss the practical repercussions of this work as well as the potential limitations it faces. The authors briefly touch upon one aspect in the final paragraph, but there would be other challenges and such an extended discussion would be critical in both highlighting the limitations as well as emphasizing the contributions of the current paper.

  Thank you for these helpful comments. We have now added a new section (the discussion section) to compare our methods with other related methods, highlight the contributions and the limitations of the paper, and also provide potential future research in detail.

Reviewer 4

Summary

This manuscript aims to eliminate the negative effects of initially poor-performing base forecasts through time series selection. In the first group of methods, based on out-of-sample information, the authors formulate the problem as an optimization problem using diverse penalty functions. The second group of methods relax the unbiasedness assumption and introduces an additional reconciliation method with selection, utilizing in-sample observations and their fitted values. Both simulation and empirical studies show the great potential of the proposed methods.

Overall, this paper provides deep insights into forecast reconciliation methods and fits well with EJOR. However, a fair bit of work is required to get published in EJOR. Specifically, the following major and minor points could be considered to improve its exposition.

Major comments

1. While the proposed methodology opts to leave *poor* base forecasts unused in the creation of reconciled forecasts, the approach by Zhang et al. (2023) is primarily focused on preserving *good* base forecasts unchanged during the reconciliation process. The key difference lies in the handling of forecasts: the former method alters the forecasts of *poor* base forecasts, ensuring these do not influence other nodes, whereas the latter method keeps the forecasts of certain nodes immutable, which then impacts others. It is crucial for the authors to emphasize these distinctions and interconnections theoretically, empirically or through discussions.

   Thank you for the suggestion. We have added a discussion section in the revised manuscript, where the first paragraph summarizes the distinctions and interconnections between our proposed methods and the approach by Zhang et al. (2023).

2. The authors' discussions on variable selection raise the question of whether they have considered a bi-level variable selection approach. Specifically, this entails allowing for both grouplevel and individual variable selection within those groups, an approach that could potentially enhance the precision and interpretability of the forecasting model. Such methodologies are well-documented in the literature, including the sparse group lasso (Simon et al., 2013), hierarchical Lasso (Zhou & Zhu, 2010), and the group bridge approach by Huang et al. (2009). Adopting a bi-level selection mechanism could provide deeper insights into the contribution of individual base forecasts, especially in terms of their significance when mapped to bottom-level disaggregated forecasts.

   Thank you for your helpful comments. In the context of forecast reconciliation, bi-level variable selection can be approached from two perspectives.

   First, following the idea of our methodology, we treat each time series in a given hierarchy as a variable. To facilitate time series selection during reconciliation, we formulate an optimization problem that controls the number of nonzero column entries in the weighting matrix G. This approach addresses individual variable selection. For group-wise variable selection, we would need to define groups of time series within the hierarchy, which is challenging due to the subjective nature of determining both the groups and the number of series within each group. Consequently, we did not explore bi-level variable selection from this perspective in our paper.

   Second, by examining the optimization problem more closely, we can treat each column of G as a group and each element as an individual. From this perspective, group-wise sparsity can be introduced by shrinking some columns of G towards zero, while within-group sparsity can be achieved by shrinking individual elements. Including an additional lasso penalty, as suggested in Simon et al. (2013), can address within-group sparsity and provide insights into the contribution of individual base forecasts when generating reconciled bottom-level forecasts. However, this would shift the focus away from our primary objective of time series selection, introduce additional hyperparameters, and increase computational complexity. Therefore, we have chosen to discuss this as a potential future research direction in the newly added discussion section.

3. Implementation of the GitHub repos provided by the authors tells that computation is somehow an issue in practice. Based on this, I suggest the authors consider the following.

- Report the computational time would provide useful guides for the readers as well as the practitioners. Give discussions in terms of complexity and scalability, especially in the context of large-scale forecasting. This is crucial for applications in real-world scenarios.

  Thank you for the comment. We have included Table 8 to present the computational time for solving the reconciliation problem for a single hierarchy in each of our experiments, considering all candidate hyperparameters and rolling test sets. We have also added a new paragraph in the discussion section to discuss the complexity and scalability of proposed methods.

- Ida, Fujiwara & Kashima (2019) proposes a fast Block Coordinate Descent for Sparse Group Lasso, which efficiently skips the updates of the groups whose parameters must be zeros by using the parameters in one group. They claim their approach reduces the processing time by up to 97% from the standard approach. I suggest the authors check whether it is helpful in improving computation.

  Thank you for the comment. We have reviewed the paper, however, they did not provide publicly available code or package to reproduce their results. Additionally, Figure 1 in Ida, Fujiwara & Kashima (2019) shows that the processing time of their method for a dataset with 74 data points and 2133 features is approximately 10^4 seconds, which is significantly longer than the times reported for a large empirical hiearchy in Table 8 of our paper. Thus, we did not consider their method in our analysis.

4. The authors claim the proposed methods keep poor base forecasts unused in generating reconciled forecasts. They should check whether this is the case at the end of their experimental studies. Specifically, an analysis should be conducted to ascertain whether the forecasts that are not selected for reconciliation indeed correspond to suboptimal ones. This could involve a detailed examination of the performance metrics of excluded forecasts compared to those included, to ensure that the selection process aligns with the stated objective of excluding poor base forecasts. Such an investigation validates the method's effectiveness and strengthens the reliability of the proposed approach in practical forecasting scenarios.

   Thank you for the helpful comment. To address this, we use the scaled error metric, RMSSE, to assess the scaled forecast variance of each series. We have now reported the RMSSE values of base forecasts for each series in both simulations, specifically in Tables 3, B.2, and B.4 for simulation setup 1, and Tables 5 and C.1-C.4 for simulation setup 2. We have also updated the related results analysis to incorporate insights from the RMSSE findings, where it was generally observed that series with higher RMSSE values are selected less frequently. For the two empirical applications, due to the large number of time series in each hierarchy, reporting RMSSE values for each series individually is impractical. Therefore, we have reported the average RMSSE values at each level to compare whether the average RMSSE for selected series is lower than that of all series at the same level. For most cases, the series selected by the proposed methods show reduced RMSSE values at each level.

   Moreover, we have included heatmaps showing the correlations between base forecast errors for both simulations. These are presented in Figures 2(b), B.1(b), and B.2(b) for setup 1, and Figures 4(b) and C.1(b) for setup 2. We have also expanded the related results analysis to incorporate insights from these results. Our analysis suggests that both forecast variance and forecast correlation influence time series selection. In general, time series with high forecast variance and strong correlations with other series are more likely to be excluded.

5. Given the variability in time series data—ranging from seasonal patterns, trend components, to noise levels—the authors could investigate how their methods perform across a diverse set of conditions. Examples are:
   - Examining the performance stability of the proposed reconciliation methods across time series with different levels of seasonality.
   - Assessing the impact of signal-to-noise ratios on the efficacy of the proposed methods.

   We agree on the importance of investigating the performance stability of the developed methods under diverse conditions. Our simulations and applications cover a broad range of data variations.

   Within a given hierarchy, we account for differences in seasonality, trend, and noise levels across various series, reflecting the inherent characteristics of hierarchical time series. For instance, the Australian labour force data shows stronger seasonality and higher signal-to-noise ratios in the most aggregated series compared to state and territory level data.

   Additionally, we address variations across different hierarchies in our paper. We generate stationary data in Section 4.2, analyze quarterly data in Section 4.1, and explore monthly data in two applications in Section 5. We also investigate the impact of correlation on reconciliation performance by controlling error correlation in the simulated hierarchy, as discussed in Section 4.2. While the simulated data in Section 4.2 is trend-free, the real-world datasets in Section 5 exhibit noticeable trends.

   Given the impracticality of covering every possible condition and the constraints of the paper's length, we have chosen to retain the experiments as presented.

6. The current methodology section and experiments provide a foundational overview of the proposed forecast reconciliation approaches and how they are implemented. However, it would benefit significantly from a more detailed exposition on several fronts to enhance the reader's understanding. Specific areas include:

   - More detailed justification for the choice of penalty functions and the theoretical underpinnings that motivated these choices.

     We have now added a paragraph in Sections 3.1.1 and 3.1.3 respectively to explain why we choose the specific penalty functions for our proposed methods, and to provide important references regarding their statistical properties.

   - The inclusion of a sensitivity analysis regarding the hyperparameters associated with each method, such as penalty parameters in the optimization problem. Understanding how variations in these parameters affect the outcomes could provide valuable insights into the robustness and flexibility of the proposed approaches.

     Thank you for the comment. We had previously outlined the parameter tuning details when introducing each proposed method in Section 3. To avoid computationally expensive cross-validation, we tune the parameters to minimize the sum of squared reconciled forecast errors on the truncated training set. This strategy is commonly used in time series forecasting. To evaluate the robustness and flexibility of the proposed approaches concerning hyperparameters, we have tested the estimated weighting matrix G derived from various candidate hyperparameter values for forecast reconciliation. The forecasting results showed significant variability in reconciliation outcomes depending on the hyperparameter values. However, the optimal parameters identified through our tuning strategy provided stable and well-performing reconciliation results. Due to space constraints, we have not included all details but have added a paragraph in Sections 4.1 and 4.2 to further explain this.

Minor comments

1. I suggest the authors put *Variable selection* as one of the keywords.

We have added "Variable selection" and removed "Grouped time series" from the keywords.

2. When cross-referencing equations, I suggest using \eqref from **amsmath** instead of \ref in the LaTeX Kernel to match the equation reference exactly.

Thanks. Now fixed.

References

Huang, J, S Ma, H Xie & CH Zhang (2009). A group bridge approach for variable selection. *Biometrika* **96**(2), 339–355.

Ida, Y, Y Fujiwara & H Kashima (2019). Fast sparse group lasso. *Advances in Neural Information Processing Systems* **32**.

Simon, N, J Friedman, T Hastie & R Tibshirani (2013). A sparse-group lasso. *J Computational and Graphical Statistics* **22**(2), 231–245.

Zhang, B, Y Kang, A Panagiotelis & F Li (2023). Optimal reconciliation with immutable forecasts. *European J Operational Research* **308**(2), 650–660.

Zhou, N & J Zhu (2010). Group variable selection via a hierarchical lasso and its oracle property. *arXiv preprint arXiv:1006.2871*.