SIS | 2022

51st Scientific Meeting of the Italian Statistical Society

Caserta, 22-24 June

Università degli Studi della Campania *Luigi Vanvitelli*

SIS Società Italiana di Statistica

www.unicampania.it

Book of the Short Papers

Editors: Antonio Balzanella, Matilde Bini, Carlo Cavicchia, Rosanna Verde

Città di Caserta

1222-2022 800 ANNI

Università degli Studi di Padova

Dipartimento di Scienze Statistiche

STATA

SAS

Università degli Studi del Sannio Benevento

Pearson

Matilde Bini (Chair of the Program Committee) - *Università Europea di Roma*
Rosanna Verde (Chair of the Local Organizing Committee) - *Università della Campania "Luigi Vanvitelli"*

PROGRAM COMMITTEE

Matilde Bini (Chair), Giovanna Boccuzzo, Antonio Canale, Maurizio Carpita, Carlo Cavicchia, Claudio Conversano, Fabio Crescenzi, Domenico De Stefano, Lara Fontanella, Ornella Giambalvo, Gabriella Grassia - Università degli Studi di Napoli Federico II, Tiziana Laureti, Caterina Liberati, Lucio Masserini, Cira Perna, Pier Francesco Perri, Elena Pirani, Gennaro Punzo, Emanuele Raffinetti, Matteo Ruggiero, Salvatore Strozza, Rosanna Verde, Donatella Vicari.

LOCAL ORGANIZING COMMITTEE

Rosanna Verde (Chair), Antonio Balzanella, Ida Camminatiello, Lelio Campanile, Stefania Capecchi, Andrea Diana, Michele Gallo, Giuseppe Giordano, Ferdinando Grillo, Mauro Iacono, Antonio Irpino, Rosaria Lombardo, Michele Mastroianni, Fabrizio Maturo, Fiammetta Marulli, Paolo Mazzocchi, Marco Menale, Giuseppe Pandolfi, Antonella Rocca, Elvira Romano, Biagio Simonetti.

ORGANIZERS OF SPECIALIZED, SOLICITED, AND GUEST SESSIONS

Arianna Agosto, Raffaele Argiento, Massimo Aria, Rossella Berni, Rosalia Castellano, Marta Catalano, Paola Cerchiello, Francesco Maria Chelli, Enrico Ciavolino, Pier Luigi Conti, Lisa Crosato, Marusca De Castris, Giovanni De Luca, Enrico Di Bella, Daniele Durante, Maria Rosaria Ferrante, Francesca Fortuna, Giuseppe Gabrielli, Stefania Galimberti, Francesca Giambona, Francesca Greselin, Elena Grimaccia, Raffaele Guetto, Rosalba Ignaccolo, Giovanna Jona Lasinio, Eugenio Lippiello, Rosaria Lombardo, Marica Manisera, Daniela Marella, Michelangelo Misuraca, Alessia Naccarato, Alessio Pollice, Giancarlo Ragozini, Giuseppe Luca Romagnoli, Alessandra Righi, Cecilia Tomassini, Arjuna Tuzzi, Simone Vantini, Agnese Vitali, Giorgia Zaccaria.

ADDITIONAL COLLABORATORS TO THE REVIEWING ACTIVITIES

Ilaria Lucrezia Amerise, Ilaria Benedetti, Andrea Bucci, Annalisa Busetta, Francesca Condino, Anthony Cossari, Paolo Carmelo Cozzucoli, Simone Di Zio, Paolo Giudici, Antonio Irpino, Fabrizio Maturo, Elvira Romano, Annalina Sarra, Alessandro Spelta, Manuela Stranges, Pasquale Valentini, Giorgia Zaccaria.

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891932310

Fully reconciled probabilistic *GDP* forecasts from Income and Expenditure sides

Riconciliazione completa delle previsioni probabilistiche del PIL dal lato del reddito e della spesa

Tommaso Di Fonzo and Daniele Girolimetto

Abstract We propose a complete reconciliation procedure of probabilistic *GDP* forecasts, resulting in *GDP* forecasts coherent with both Income and Expenditure sides' forecasted series, and evaluate its performance on the Australian quarterly *GDP* series, as compared to the original proposal by Athanasopoulos *et al.* (2020).
Abstract *In questo lavoro viene proposta una procedura di riconciliazione delle previsioni probabilistiche del PIL e delle sue componenti tanto dal lato del Reddito quanto da quello della Spesa, volta a produrre previsioni coerenti rispetto ad entrambi i lati. Tale procedura, applicata alle serie trimestrali del PIL australiano, viene posta a confronto con la proposta originale di Athanasopoulos et al. (2020).*

Key words: probabilistic forecast reconciliation, linearly constrained multiple time series, GDP, Income, Expenditure

1 Introduction and summary

Probabilistic forecasting has become relevant and widely used in many fields in recent years (Gneiting and Katzfuss, 2014). Panagiotelis *et al.* (2020) propose an effective definition of reconciled (coherent) probabilistic forecasts in the case of genuine[1] hierarchical/grouped time series, that has been further worked out by Wickramasuriya (2021) in a Gaussian framework.

The reconciliation of point and probabilistic forecasts of Australian *GDP* from Income and Expenditure sides, each one distinctly dealt with, was originally considered by Athanasopoulos *et al.* (2020) in order to perform aligned decision making and to improve forecast accuracy. In their empirical study they consider 95 Australian Quarterly National Accounts time series, describing the Gross Domestic Product (*GDP*) at current prices from Income and Expenditure sides, interpreted as two distinct hierarchical structures. In the former case (Income), *GDP* is on the top of 15 lower level aggregates, while in the latter (Expenditure), *GDP* is the top level aggregate of a hierarchy of 79 time series (for details, see Athanasopoulos *et al.*, 2020, and Bisaglia *et al.*, 2020).

T. Di Fonzo (e-mail: tommaso.difonzo@unipd.it); D. Girolimetto (e-mail: daniele.girolimetto@phd.unipd.it); Department of Statistical Sciences, University of Padova.

[1] In a genuine hierarchical/grouped time series, the link between the top-level and $n_b > 1$ bottom variables is uniquely defined (Di Fonzo and Girolimetto, 2021).

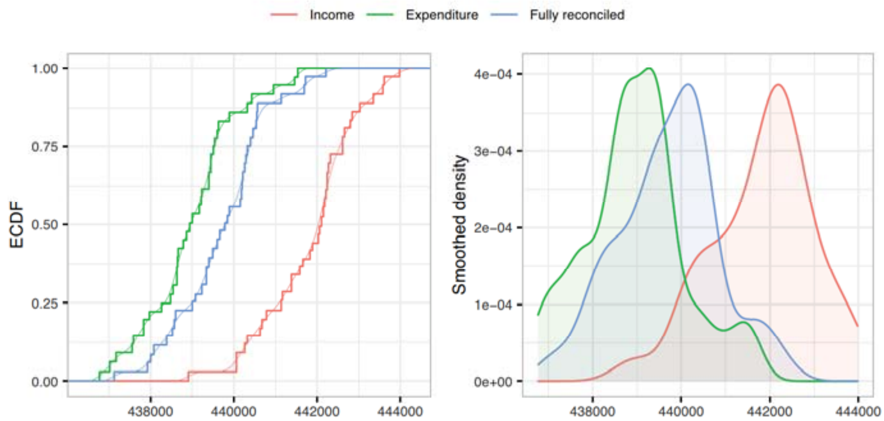

Fig. 1 *GDP* empirical 1-step-ahead forecast distributions for 2018:Q1, MinT-shr reconciliation approach. Empirical Cumulative Distribution Function (left), and Smoothed density (right).

In this paper, the results shown in Bisaglia *et al.* (2020) as for the reconciliation of point macroeconomic forecasts, are extended to a probabilistic forecasting framework. We re-consider the results of Athanasopoulos *et al.* (2020), where the probabilistic forecasts of the Australian quarterly *GDP* aggregates are separately reconciled from Income ($\widetilde{GDP}^I$) and Expenditure ($\widetilde{GDP}^E$) sides. This means that the empirical forecast distributions $\widetilde{GDP}^I$ and $\widetilde{GDP}^E$ are each coherent (according to Theorem 3.5 in Panagiotelis *et al.*, 2020) within its own pertaining side with the other empirical forecast distributions, but in general $\widetilde{GDP}^I \neq \widetilde{GDP}^E$ at any forecast horizon. This circumstance could confuse the user, mostly when the difference between the empirical forecast distributions is not negligible, as shown in Figure 1, where the *GDP* empirical forecast distributions from Income and Expenditure sides for 2018:Q1 are presented along with their fully reconciled counterparts through the MinT-shr approach (Athanasopoulos *et al.*, 2020). Bisaglia *et al.* (2020) propose a complete reconciliation strategy, able to produce a 'one number point forecast' of the *GDP* figure. However, a similar result is still missing for the probabilistic reconciliation setting. We fill this gap, showing how (incoherent) base probabilistic forecasts for general linearly constrained multiple time series (i.e., not necessarily genuine hierarchical/grouped time series) may be coherently reconciled[2].

2 Coherent probabilistic forecast reconciliation

We consider a linearly constrained n-dimensional multiple time series $\{\mathbf{y}_t\}_{t=1}^T$, with constraints expressed in homogeneous form as $\mathbf{U}'\mathbf{y}_t = \mathbf{0}_{r\times 1}$, where $\mathbf{U}'$ is a $(r \times n)$ matrix of known coefficients. For a genuine hierarchical/grouped time series, setting $n_a = r$, the standard structural representation (Hyndman *et al.*, 2011) holds:

[2] Note that the naive practice of averaging *GDP* forecasts from different sides yields a single forecast, that is though inconsistent with the component variables from both sides.

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t, \quad t = 1, \ldots, T, \tag{1}$$

where $\mathbf{S} = \begin{bmatrix}\mathbf{C}' & \mathbf{I}_{n_b}\end{bmatrix}'$ is the summing matrix mapping the n_b bottom time series $\mathbf{b}_t$ into the complete vector $\mathbf{y}_t = \begin{bmatrix}\mathbf{a}_t' & \mathbf{b}_t'\end{bmatrix}'$. The $(n_a \times 1)$ vector $\mathbf{a}_t$ contains the aggregated (upper) time series, $\mathbf{C} \in \{0,1\}^{n_a \times n_b}$ is the contemporaneous aggregation matrix (Di Fonzo and Girolimetto, 2021), $\mathbf{a}_t = \mathbf{C}\mathbf{b}_t$, and $\mathbf{U}' = \begin{bmatrix}\mathbf{I}_{n_a} & -\mathbf{C}\end{bmatrix}$.

To exploit definitions and results of Panagiotelis *et al.* (2020), valid for genuine hierarchical/grouped time series expressed as in (1), we need to find out a relationship like expression (1) for a general linearly constrained multiple time series as well. In general, for example when the number of constraints makes manual handling of the problem difficult, this can be done in the following steps:

1. compute $\mathbf{R}'$, the $(n_v \times n_f)$ reduced row echelon form (*rref*, Strang, 2019) of matrix $\mathbf{U}'$, where possible $r - n_v$ zero rows are not considered;
2. find out a $(n \times n)$ permutation matrix $\bar{\mathbf{P}}$, such that $\bar{\mathbf{U}}' = \mathbf{R}'\bar{\mathbf{P}}'$ has the form $\bar{\mathbf{U}}' = \begin{bmatrix}\mathbf{I}_{n_v} & -\bar{\mathbf{C}}\end{bmatrix}$, where $\bar{\mathbf{C}}$ is a $(n_v \times n_f)$ matrix of known coefficients (differently from matrix $\mathbf{C}$, the items of matrix $\bar{\mathbf{C}}$ are not necessarily equal to either 0 or 1, and may be negative);
3. define the matrix $\bar{\mathbf{S}} = \begin{bmatrix}\bar{\mathbf{C}}' & \mathbf{I}_{n_f}\end{bmatrix}'$ (analogous of the summing matrix $\mathbf{S}$), and $\bar{\mathbf{y}}_t = \bar{\mathbf{P}}\mathbf{y}_t = \begin{bmatrix}\mathbf{v}_t' & \mathbf{f}_t'\end{bmatrix}'$, where $\mathbf{f}_t$ is a $(n_f \times 1)$ vector of 'free' variables, and $\mathbf{v}_t$ is a $(n_v \times 1)$ vector of 'basic' variables, such that $\mathbf{v}_t = \bar{\mathbf{C}}\mathbf{f}_t$, $t = 1, \ldots, T$;
4. define the 'structural' representation of a general linearly constrained multiple time series as
$$\bar{\mathbf{y}}_t = \bar{\mathbf{S}}\mathbf{f}_t, \quad t = 1, \ldots, T; \tag{2}$$
5. consider the general optimal combination reconciliation formula according to the projection approach (Di Fonzo and Girolimetto, 2021), $\tilde{\mathbf{y}}_h = \mathbf{M}\hat{\mathbf{y}}_h$, where $\mathbf{M} = \mathbf{I}_n - \mathbf{W}_h\mathbf{U}\left(\mathbf{U}'\mathbf{W}_h\mathbf{U}\right)^{-1}\mathbf{U}'$, with $\mathbf{W}_h$ $(n \times n)$ p.d. matrix, and express the same result as:
$$\bar{\mathbf{P}}\tilde{\mathbf{y}}_h = \bar{\mathbf{S}}\bar{\mathbf{G}}\bar{\mathbf{P}}\hat{\mathbf{y}}_h, \tag{3}$$
where $\bar{\mathbf{G}} = \left(\bar{\mathbf{S}}'\bar{\mathbf{P}}'\mathbf{W}_h^{-1}\bar{\mathbf{P}}\bar{\mathbf{S}}\right)^{-1}\bar{\mathbf{S}}'\bar{\mathbf{P}}'\mathbf{W}_h^{-1}$;
6. given n_f (and $n_v = n - n_f$), $\tilde{\mathbf{y}}_h$ is uniquely determined and does not depend on a particular choice of the free variables. This point is formally clarified in the extended version of the paper.

We consider 3 cases for the approximation of the covariance matrix $\mathbf{W}_h$:

- ols: $\mathbf{W}_h = \sigma^2\mathbf{I}_n$ (Hyndman *et al.*, 2011),
- wls: $\mathbf{W}_h = \widehat{\mathbf{W}}_D = \text{diag}\,\widehat{\mathbf{W}}_1$ (Hyndman *et al.*, 2016),
- shr: $\mathbf{W}_h = \widehat{\mathbf{W}}_{shr} = \lambda\widehat{\mathbf{W}}_D + (1-\lambda)\widehat{\mathbf{W}}_1$ (Wickramasuriya *et al.*, 2019),

where $\widehat{\mathbf{W}}_1$ is the $(n \times n)$ covariance matrix of the in-sample one-step-ahead base forecasts errors, and $\widehat{\mathbf{W}}_{shr}$ is its shrinked version (Wickramasuriya *et al.*, 2019).

Representation (2) means that $\bar{\mathbf{y}}_t$ lies in an n-dimensional subspace of $\mathbb{R}^n$ spanned by the columns of $\bar{\mathbf{S}}$, called "coherent subspace" and denoted by $\overline{\mathscr{S}}$ (Panagiotelis *et al.*, 2020). Now, let $\mathscr{F}_{\mathbb{R}^{n_f}}$ be the Borel σ-algebra on $\mathbb{R}^{n_f}$, $\left(\mathbb{R}^{n_f}, \mathscr{F}_{\mathbb{R}^{n_f}}, \nu\right)$ a probability space for the free variables, and $\bar{s}: \mathbb{R}^{n_f} \rightarrow \mathbb{R}^n$ a continuous mapping matrix. Then a σ-algebra $\mathscr{F}_{\overline{\mathscr{S}}}$ can be constructed from the collection of sets $\bar{s}(\mathscr{B})$ for all $\mathscr{B} \in \mathscr{F}_{\mathbb{R}^n}$.

Definition 1. (*Coherent probabilistic forecast for a linearly constrained multiple time series*) Given the triple $\left(\mathbb{R}^{n_f}, \mathscr{F}_{\mathbb{R}^{n_f}}, \nu\right)$, we define a coherent probability triple $\left(\overline{\mathscr{S}}, \mathscr{F}_{\overline{\mathscr{S}}}, \breve{\nu}\right)$ such that $\breve{\nu}(\bar{s}(\mathscr{B})) = \nu(\mathscr{B})$, $\forall \mathscr{B} \in \mathscr{F}_{\mathbb{R}^{n_f}}$.

In order to extend forecast reconciliation to the probabilistic setting, let $(\mathbb{R}^n, \mathscr{F}_{\mathbb{R}^n}, \widehat{\nu})$ be a probability triple characterizing base (incoherent) probabilistic forecasts for all n series, and let $\psi : \mathbb{R}^{n_f} \rightarrow \mathbb{R}^n$ be a continuous mapping function defined by Panagiotelis *et al.* (2020) as the composition of two transformations, $\bar{s} \circ \bar{g}$, where $\bar{g} : \mathbb{R}^n \rightarrow \mathbb{R}^{n_f}$ is a continuous function corresponding to matrix $\overline{\mathbf{G}}$ in equation (3).

Definition 2. (*Probabilistic forecast reconciliation for a linearly constrained multiple time series*). The reconciled probability measure of $\hat{\nu}$ with respect to ψ is a probability measure $\tilde{\nu}$ on $\overline{\mathscr{S}}$ with σ-algebra $\mathscr{F}_{\overline{\mathscr{S}}}$ such that

$$\tilde{\nu}(\mathscr{A}) = \hat{\nu}\left(\psi^{-1}(\mathscr{A})\right), \quad \forall \mathscr{A} \in \mathscr{F}_{\overline{\mathscr{S}}}, \tag{4}$$

where $\psi^{-1}(\mathscr{A}) = \{x \in \mathbb{R}^n : \psi(x) \in \mathscr{A}\}$ is the pre-image of $\mathscr{A}$.

3 Joint bootstrap-based probabilistic forecast reconciliation

Since an analytical expression of the forecast distribution is either unavailable, or relies on unrealistic parametric assumptions, the empirical evaluation of the results will be grounded on reconciled samples obtained according to Theorem 3.5 in Panagiotelis *et al.* (2020)[3]:

Theorem 1. *(Reconciled samples). Suppose that* $\left(\hat{\mathbf{y}}^{[1]}, \ldots, \hat{\mathbf{y}}^{[L]}\right)$ *is a sample drawn from an incoherent probability measure* $\hat{\nu}$*. Then* $\left(\tilde{\mathbf{y}}^{[1]}, \ldots, \tilde{\mathbf{y}}^{[L]}\right)$*, where* $\tilde{\mathbf{y}}^{[\ell]} := \psi\left(\hat{\mathbf{y}}^{[\ell]}\right)$ *for* $\ell = 1, \ldots, L$*, is a sample drawn from the reconciled probability measure* $\tilde{\nu}$ *as defined in (4).*

According to this theorem, reconciling each member of a sample obtained from an incoherent distribution yields a sample from the reconciled distribution. As a consequence, coherent probabilistic forecasts may be developed through a post-forecasting mechanism analogous to the point forecast reconciliation setting. For this purpose, the bootstrap procedure by Gamakumara *et al.* (2018) is applied:

1. appropriate univariate models M_i for each series in the system are fitted based on the training data $\{y_{i,t}\}_{t=1}^T$, $i = 1, \ldots, n$, and the one-step-ahead in-sample forecast errors are stacked in an $(n \times T)$ matrix, $\widehat{\mathbf{E}} = \{\hat{e}_{i,t}\}$;
2. $\hat{\mathbf{y}}_{i,h}^{[l]} = f_i\left(M_i, \hat{e}_{i,h}^{[l]}\right)$ is computed for $h = 1, \ldots, H$ and $l = 1, \ldots, L$, where $f(\cdot)$ is a function of the fitted univariate model and associated error, $\hat{\mathbf{y}}_{i,h}^{[l]}$ is a sample path simulated for the $i-$th series, and $\hat{e}_{i,h}^{[l]}$ is the $(i,h)-$th element of an $(n \times H)$ block bootstrap matrix containing H consecutive columns randomly drawn from $\hat{\mathbf{E}}$;
3. the optimal combination reconciliation formula (3) is performed for each $\hat{\mathbf{y}}_h^{[l]}$.

[3] Extension to a linearly constrained multiple time series for the Gaussian case (Wickramasuriya, 2021) is currently under study, and will be presented in the extended version of the paper.

The accuracy of the probabilistic forecasts is evaluated using the Cumulative Rank Probability Score (CRPS, Gneiting and Katzfuss, 2014, Panagiotelis *et al.*, 2020). In addition, we employ the Energy Score (ES), that is the CRPS extension to the multivariate case, to evaluate the forecasting accuracy for the whole system (Gamakumara *et al.*, 2018).

4 Reconciled probabilistic forecasts of the Australian *GDP*

For the complete Australian *GDP* accounts from both Income and Expenditure sides, it is $n = 95$ ($n_f = 62$ and $n_v = 33$), and the homogeneous constraints are described by matrix $\mathbf{U}'$ shown in Bisaglia *et al.* (2020). In addition, the available time series span over the period 1984:Q4 - 2018:Q1.

Athanasopoulos *et al.* (2020) produced base forecasts for each time series using univariate ARIMA models (with the `auto.arima` function of the R-package). Our reconciliation proposal is applied within the same forecasting experiment, that considers forecasts from $h = 1$ quarter ahead up to $h = 4$ quarters ahead, using an *expanding* window, where the first training sample is set from 1984:Q4 to 1994:Q3. The base forecasts are reconciled using the `htsrec` function of the R-package `FoReco` (Girolimetto and Di Fonzo, 2022).

Figure 2 shows the *skill scores* of CRPS for the *GDP* reconciled forecasts (top panel), and of ES for Income and Expenditure sides, and the whole system (bottom panel), that is the percentage changes registered by these indices for the considered reconciliation procedures, computed such that positive values signal an improvement in forecasting accuracy over the base forecasts. 'Income' and 'Expenditure' panels, respectively, refer to the results found by Athanasopoulos *et al.* (2020), while the 'Fully reconciled' panel shows the *skill scores* of the fully reconciled probabilistic forecasts. From the bottom panel of Figure 2, it appears that the reconciliation improves on the base forecasts' accuracy, shr and wls offering almost always the best performance. For *GDP*, ols outperforms both wls and shr, whatever side is considered. In general, the simultaneously reconciled probabilistic forecasts give results as good as those of Athanasopoulos *et al.* (2020). In addition, the newly proposed approach produces forecasts that are fully coherent with all economic constraints coming from National Accounts relationships.

References

1. Athanasopoulos, G., Gamakumara, P., Panagiotelis, A., Hyndman, R.J., Affan, M. (2020), Hierarchical Forecasting, in Fuleky, P. (ed.), *Macroeconomic Forecasting in the Era of Big Data*, Cham, Springer, 689–719.
2. Bisaglia, L., Di Fonzo, T., Girolimetto, D. (2020), Fully reconciled GDP forecasts from Income and Expenditure sides, in Pollice, A., Salvati, N., Schirripa Spagnolo F. (eds.), *Book of short papers SIS 2020*, 951–956, Pearson.
3. Di Fonzo, T., and Girolimetto, D. (2021), Cross-temporal forecast reconciliation: Optimal combination method and heuristic alternatives, *International Journal of Forecasting*, in press.
4. Gamakumara, P., Panagiotelis, A., Athanasopoulos, G., Hyndman, R.J. (2018), *Probabilistic Forecasts in Hierarchical Time Series*, Department of Econometrics and Business Statistics, Monash University, Working Paper 11/18.
5. Girolimetto, D., Di Fonzo, T. (2022), *FoReco: Point Forecast Reconciliation*, R package version 0.2.4, `https://CRAN.R-project.org/package=FoReco`.
6. Gneiting, T. and Katzfuss, M. (2014), Probabilistic forecasts, *Annual Review of Statistics and Its Application*, 1, 125-151.

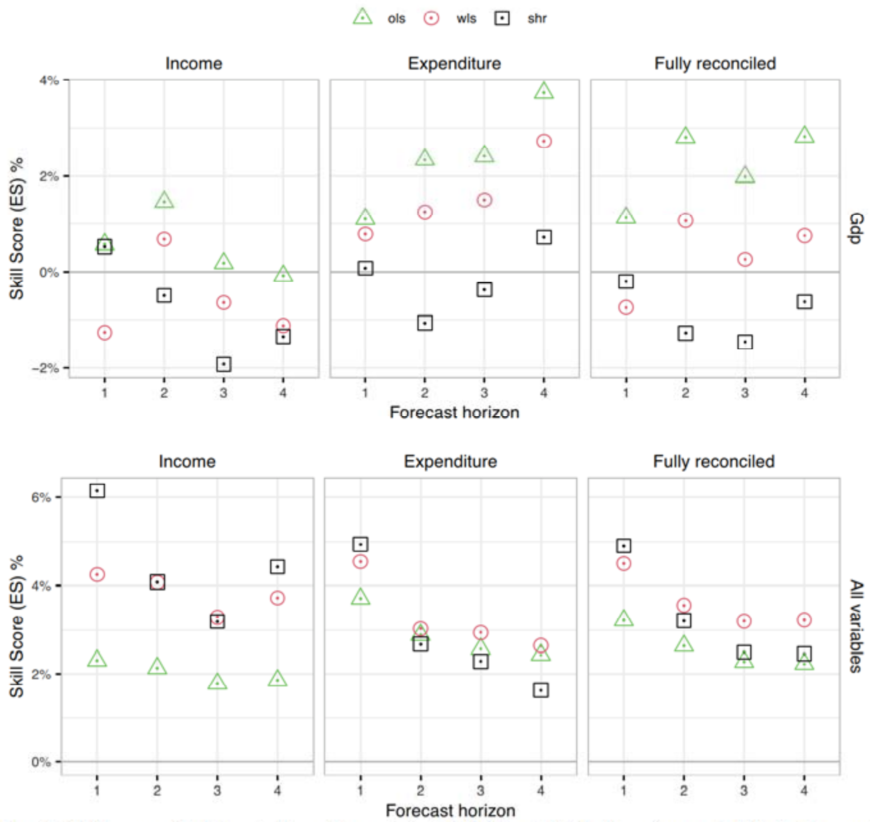

Fig. 2 Skill scores (relative to base forecasts) of CRPS and ES indices for probabilistic forecasts from alternative reconciliation approaches.

7. Hyndman, R.J., Ahmed, R.A., Athanasopoulos, G., Shang, H.L. (2011), Optimal combination forecasts for hierarchical time series, *Computational Statistics and Data Analysis*, 55, 2579–2589.
8. Hyndman, R. J., Lee, A.J., and Wang E., (2016), Fast computation of reconciled forecasts for hierarchical and grouped time series, *Computational Statistics and Data Analysis*, 97, 16–32.
9. Panagiotelis, A., Gamakumara, P., Athanasopoulos, G., Hyndman, R.J. (2020), *Probabilistic Forecast Reconciliation: Properties, Evaluation and Score Optimisation*, Monash University, Department of Econometrics and Business Statistics, Working Paper 26/20.
10. Panagiotelis, A., Gamakumara, P., Athanasopoulos, G., Hyndman, R.J. (2021), Forecast reconciliation: A geometric view with new insights on bias correction, *International Journal of Forecasting*, 37, 1, 343–359.
11. Strang, G., (2019), *Linear algebra and learning from data*, Wellesley, Cambridge Press.
12. Wickramasuriya, S.L., (2021), Probabilistic forecast reconciliation under the Gaussian framework, `arXiv.2103.11128`.
13. Wickramasuriya, S.L., Athanasopoulos, G., 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.

Acknowledgments

The authors acknowledge financial support from project PRIN2017 "HiDEA: Advanced Econometrics for High-frequency Data", 2017RSMPZZ.