Article

An Alternative Proof of Minimum Trace Reconciliation

Sakai Ando * and **Futoshi Narita**

Research Department, International Monetary Fund, 700 19th St NW, Washington, DC 20431, USA;
fnarita@imf.org
* Correspondence: sando@imf.org

Abstract: Minimum trace reconciliation, developed by Wickramasuriya et al., 2019, is an innovation in the literature on forecast reconciliation. The proof, however, has a gap, and the idea is not easy to extend to more general situations. This paper fills the gap by providing an alternative proof based on the first-order condition in the space of a non-square matrix and arguing that it is not only simpler but also can be extended to incorporate more general results on minimum weighted trace reconciliation in Panagiotelis et al., 2021. Thus, our alternative proof not only has pedagogical value but also connects the results in the literature from a unified perspective.

Keywords: forecast reconciliation; minimum trace reconciliation; hierarchical time series

JEL Classification: C53; E17

1. Introduction

Minimum trace reconciliation, developed by [1], is an innovation in the literature on forecast reconciliation. The tool enables a systematic approach to forecasting with linear constraints, which encompasses a wide range of applications, including electricity demand forecasting [2] and macroframework forecasting [3,4].

The proof of [1], however, has a gap and is not easy to extend to more general situations. Their proof attempts to solve a minimization problem by replacing the objective function with its lower bound. Although they find the solution that minimizes the lower bound, the minimizer is not shown to coincide with the solution to the original problem, which creates a gap in the proof.

This paper provides an alternative proof and argues that it is not only simpler but also can be extended to incorporate more general results in the literature. The proof is more direct in the sense that it solves the first-order condition in the space of the non-square matrix. An almost identical proof can be used to prove Theorem 3.3 of [5], which shows that the minimum trace reconciliation and minimum weighted trace reconciliation lead to an identical formula. By selecting a special weight in the weighted trace reconciliation problem, we can also see why the lower bound minimization in [1] reaches the same formula. Thus, the alternative proof not only has pedagogical value but also connects the results in the literature from a unified perspective.

The paper is organized into six sections. In Section 2, we provide the setup of the problem. In Section 3, we briefly illustrate the proof of [1]. In Section 4, we provide an alternative proof of [1]. Section 5 extends the proof to incorporate [5] and discusses the insights. In Section 6, we make our conclusions.

2. Setup

The setup and notation follow [1]. Let y_t and b_t be $m \times 1$ and $n \times 1$ vectors of random variables, where $m > n > 0$. The two vectors are constrained linearly by

$$y_t = Sb_t, \tag{1}$$

where S is a $m \times n$ matrix, and its last n rows are identity matrix

$$S = \begin{bmatrix} C \\ I_n \end{bmatrix}, \tag{2}$$

and, thus, S is of full column rank for any matrix C. Intuitively, b_t represents the most disaggregated level and y_t includes b_t itself and aggregates of the subcomponents as specified by C, although mathematically, C can include negative elements. In any case, the realization of y_t is linearly dependent and belongs to

$$\mathcal{A} = \{y \in \mathbb{R}^m : [I_{m-n} - C]y = 0\} \tag{3}$$

as $\begin{bmatrix} I_{m-n} & -C \end{bmatrix} \begin{bmatrix} C \\ I_n \end{bmatrix} = C - C = 0.$

Suppose that an h-step ahead forecast based on the information up to time T, denoted by $\hat{y}_T(h)$ and called "base" forecast, is given. The base forecast $\hat{y}_T(h)$ is assumed to be an unbiased estimator of y_{T+h}

$$\mathbb{E}_T y_{T+h} = \mathbb{E}_T \hat{y}_T(h), \tag{4}$$

where $\mathbb{E}_T$ is the expectation conditional on the information up to time T. But an issue is that $\hat{y}_T(h)$ may not belong to $\mathcal{A}$, which motivates forecast reconciliation.

A reconciled forecast $\widetilde{y}_T(h)$ given an $n \times m$ matrix P is a linear transformation of $\hat{y}_T(h)$ such that

$$\widetilde{y}_T(h) = SP\hat{y}_T(h). \tag{5}$$

The role of P is to map the base forecast $\hat{y}_T(h)$ into the most disaggregated level. The reconciled forecast $\widetilde{y}_T(h)$ is assumed to be unbiased and, thus, satisfies

$$\begin{aligned} &\mathbb{E}_T \widetilde{y}_T(h) = \mathbb{E}_T y_{T+h} = \mathbb{E}_T \hat{y}_T(h), \;\; \forall y_{T+h} \\ \Leftrightarrow\; &SPS\mathbb{E}_T b_{T+h} = S\mathbb{E}_T b_{T+h}, \;\; \forall \mathbb{E}_T b_{T+h} \\ \Leftrightarrow\; &SPS = S \\ \Leftrightarrow\; &PS = I_n. \end{aligned} \tag{6}$$

Note that the necessity of the last equivalence follows from multiplying S' from the left of both sides. $S'S$ is a full-rank square matrix as S is a full-rank matrix, so $S'S$ is invertible. The sufficiency follows from multiplying S from the left of both sides.

The forecast error of the reconciled forecast can be expressed as

$$\mathbb{E}_T \left[\left(y_{T+h} - \widetilde{y}_T(h) \right) \left(y_{T+h} - \widetilde{y}_T(h) \right)' \right] = SPWP'S', \tag{7}$$

where $W = \mathbb{E}_T \left[(y_{T+h} - \hat{y}_T(h))(y_{T+h} - \hat{y}_T(h))' \right]$ is the covariance matrix of the h-step ahead base forecast error and is assumed to be invertible (i.e., excluding the case of zero forecast error and the case of degenerated matrix C for aggregation). The equality follows because

$$y_{T+h} - \widetilde{y}_T(h) = y_{T+h} - SP\hat{y}_T(h) = \underbrace{(I_m - SP)S}_{=0}b_{T+h} + SP(y_{T+h} - \hat{y}_T(h)), \tag{8}$$

noting that $(I_m - SP)S = S - SPS = 0$ holds. Ref. [1] attempted to prove that the matrix P that minimizes the trace of the covariance matrix subject to the unbiasedness constraint is

$$\left(S'W^{-1}S\right)^{-1}S'W^{-1} = arg \min_{P \in \mathbb{R}^{n \times m}} tr\left[SPWP'S'\right] \; s.t. \; PS = I_n. \tag{9}$$

3. Gap in Proof of [1]

The proof of [1] can be divided into two steps. First, they show in its online Appendix A2 that the objective function can be bounded from below, as follows:

$$tr\left[SPWP'S'\right] \geq tr\left[PWP'\right], \;\; \forall P \in \mathbb{R}^{n \times m}. \tag{10}$$

Second, a minimization problem where the objective function is the lower bound is solved.

$$\left(S'W^{-1}S\right)^{-1}S'W^{-1} = arg\min_{P \in \mathbb{R}^{n \times m}} tr\left(PWP'\right) \ s.t. \ PS = I_n. \tag{11}$$

The proof ends here, and thus, one still needs to show that the minimizers of the two problems (9) and (11) coincide. This presents a gap in the proof.

This gap is non-trivial because minimizing a function is not generally the same as minimizing its lower bound. That is, a function $f(x)$ being bounded from below by another function $g(x)$ does not guarantee that their minimizers coincide. For a counterexample where minimizers do not coincide, consider

$$f(x) = 2x^2 - 4x + 4, \ g(x) = x^2, \ x \in \mathbb{R}, \tag{12}$$

where $f(x)$ is bounded by $g(x)$ from below, as follows:

$$f(x) - g(x) = (x-2)^2 \geq 0, \forall x \in \mathbb{R}, \tag{13}$$

but the minimizers do not coincide, as follows:

$$\underset{x \in \mathbb{R}}{argmin} f(x) = 1 \neq \underset{x \in \mathbb{R}}{argmin} g(x) = 0. \tag{14}$$

In the case of (9) and (11), however, the minimizers do coincide, as explained in proposition 2 of Section 5.

4. An Alternative Proof of (9)

The alternative proof that we propose is an extension of the partial derivative and the first-order condition in a space of the matrix.

Proof. Let $(\mathbb{R}^{n \times m}, \langle, \rangle)$ be the space of $n \times m$ matrix equipped with the Frobenius inner product ([6]):

$$\langle A, B \rangle = tr\left(A'B\right), \ A, B \in \mathbb{R}^{n \times m}. \tag{15}$$

By Theorem 1 of [7] (p. 243), there exists an $n \times m$ matrix Lagrange multiplier Λ such that the Lagrangian

$$L(P) = tr\left(SPWP'S'\right) + tr\left(\Lambda'(I_n - PS)\right) \tag{16}$$

is stationary at its minimum point. This means that, at the minimum, the directional derivative (or Gateaux differential, as defined on page 171 of [7]) of $L(P)$ is zero for any $n \times m$ matrix H.

By a direction calculation,

$$
\begin{aligned}
L(P + \alpha H) &= tr\left(S(P+\alpha H)W(P+\alpha H)'S'\right) + tr\left(\Lambda'(I_n - (P+\alpha H)S)\right) \\
&= tr(SPWP'S') + tr\left(\Lambda'(I_n - PS)\right) + \alpha tr(SHWPS') + \alpha tr(SPWH'S') \\
&\quad + \alpha^2 tr(SHWH'S') - \alpha tr\left(\Lambda'HS\right) \\
&= L(P) + \alpha tr(SHWPS') + \alpha tr(SPWH'S') + \alpha^2 tr(SHWH'S') - \alpha tr\left(\Lambda'HS\right).
\end{aligned}
\tag{17}
$$

By rearranging the terms and taking the limit in α, the quadratic term disappears, and the derivative becomes the following:

$$
\begin{aligned}
\lim_{\alpha \to 0} \frac{L(P+\alpha H) - L(P)}{\alpha} &= tr\left(SHWP'S' + SPWH'S' - \Lambda'HS\right) \\
&= tr\left(H\left(2WP'S'S - S\Lambda'\right)\right) \\
&= 0,
\end{aligned}
\tag{18}
$$

where the second equality uses $tr(A'B) = tr(B'A) = tr(BA')$ and the symmetry of W. Since this has to hold for all H,

$$2WP'S'S = S\Lambda' \Rightarrow P'S'S = \frac{1}{2}W^{-1}S\Lambda'. \tag{19}$$

Multiplying S' on both sides from the left and using $SPS = S$ gives the following:

$$S'S = \frac{1}{2} S' W^{-1} S \Lambda' \Rightarrow \left(S' W^{-1} S \right)^{-1} S'S = \frac{1}{2} \Lambda'. \tag{20}$$

Thus, the formula is as follows:

$$P'S'S = W^{-1} S \left(S' W^{-1} S \right)^{-1} S'S \Rightarrow P = \left(S' W^{-1} S \right)^{-1} S' W^{-1}. \tag{21}$$

$\square$

The proof essentially uses the extension of a partial derivative and solves the first-order condition. Since the objective function is quadratic and the constraint is linear, the first-order condition is sufficient.

5. An Extension of the Alternative Proof

The proof can be applied to the environment of weighted trace minimization as Theorem 3.3 of [5]. To motivate the extension, suppose we have the base forecast of the variables in the GDP expenditure approach and want to reconcile it to satisfy

$$Y = C + I + G + XM, \tag{22}$$

where Y is GDP, C is consumption, I is investment, G is government expenditure, and XM is net export. The minimum trace reconciliation minimizes the variance of forecast error with equal weights, as follows:

$$V(Y - \widetilde{Y}) + V(C - \widetilde{C}) + V(I - \widetilde{I}) + V(G - \widetilde{G}) + V(XM - \widetilde{XM}) \tag{23}$$

subject to the constraint (22). Since the forecast of GDP often attracts more attention than the others, a natural question is whether it is possible to improve the forecast of some variables at the expense of other variables by adjusting the weights.

$$\omega_Y V(Y - \widetilde{Y}) + \omega_C V(C - \widetilde{C}) + \omega_I V(I - \widetilde{I}) + \omega_G V(G - \widetilde{G}) + \omega_{XM} V(XM - \widetilde{XM}), \ \omega_i > 0, \ \sum_i \omega_i = 1. \tag{24}$$

Such specification can be expressed as a weighted trace, as follows:

$$\min_{P \in \mathbb{R}^{n \times m}} tr \left(\omega SPWP'S' \right) \ s.t. \ PS = I_n, \tag{25}$$

where ω is a $m \times m$ matrix with its (i, i) element equal to ω_i. When the $m \times m$ weight matrix is a diagonal matrix, the objective function is a weighted sum of the variance of forecast errors. Note that the constraint in (25) is the same as that in (9), so the same unbiasedness assumption is still imposed. The same unbiasedness assumption is also reflected in the objective function as in (9).

As [5] showed and [1] proved in their unpublished manuscript, the optimal matrix P is independent of ω as long as ω is symmetric and invertible. Therefore, in practice, one does not need to exercise judgment or estimate how much weight to put on which variable.

Proposition 1. *For any symmetric and invertible $m \times m$ matrix ω, the solution to (25) is*

$$P = (S' W^{-1} S)^{-1} S' W^{-1}. \tag{26}$$

Proof. The proof is almost identical to Section 4. Let the Lagrangian be

$$L(P) = tr \left(\omega SPWP'S' \right) + tr \left(\Lambda'(I_n - PS) \right). \tag{27}$$

Following the same logic as Section 4, the first-order condition leads to

$$\lim_{\alpha \to 0} \frac{L(P + \alpha H) - L(P)}{\alpha} = tr \left(H (2WP'S' \omega S - S\Lambda') \right) = 0. \tag{28}$$

Since this has to hold for all H,

$$2WP'S'\omega S = S\Lambda' \Rightarrow P'S'\omega S = \frac{1}{2}W^{-1}S\Lambda'. \tag{29}$$

Multiplying S' on both sides from the left and using $SPS = S$ gives the following:

$$S'\omega S = \frac{1}{2}S'W^{-1}S\Lambda' \Rightarrow \left(S'W^{-1}S\right)^{-1}S'\omega S = \frac{1}{2}\Lambda'. \tag{30}$$

The formula follows because $S'\omega S$ is a full-rank square matrix and, thus, invertible.

$$P'S'\omega S = W^{-1}S\left(S'W^{-1}S\right)^{-1}S'\omega S \Rightarrow P = \left(S'W^{-1}S\right)^{-1}S'W^{-1}. \tag{31}$$

$\square$

Intuitively, the fact that the weight matrix does not matter can be interpreted as saying that there is not a trade-off between variables, as if the choice matrix P has enough degree of freedom in mixing the base forecast so that the variance of each variable's forecast error can be minimized variable by variable, without affecting the variance of other variables' forecast errors.

Mathematically, the proof shares an almost identical structure as Section 4, which is a special case when $\omega = I_m$. Since a symmetric invertible matrix can be factorized as $\omega = A'A$ from Takagi's factorization, the objective function can be written as

$$tr\left(\omega SPWP'S'\right) = tr\left(ASPWP'S'A'\right) \tag{32}$$

for any full-rank square matrix A. In fact, since the proof only requires $S'\omega S$ to be invertible, one can extend A to be a non-square matrix and show that the objective function of (11) is a special case of that of (25).

Proposition 2. *There exists a weight ω such that the objective function of (11) equals that of (25).*

Proof. Let $A := (S'S)^{-1}S'$ and $\omega := A'A$. The objective function of (25) collapses to that of (11).

$$tr\left(\omega SPWP'S'\right) = tr\left(ASPWP'S'A'\right) = tr\left(PWP'\right). \tag{33}$$

$\square$

Note that since $S'\omega S = I_n$ is invertible, the proof of proposition 1 can be applied to show (11). This is one way to see why the proof of [1] reaches the same formula. One insight from the right side of (33) is that it represents the summed variance of the forecast error of the most disaggregated variables. Thus, minimizing the summed variance of all variables is equivalent to minimizing the summed variance of the most disaggregated variables.

In summary, the extension to allow a general weight highlights two observations. First, the irrelevance of weight implies that the objective function, being the trace of the forecast error covariance matrix, is not essential, although (9) is called minimum trace reconciliation in the literature. What is essential is the unbiasedness assumption, and thus, it could alternatively be called an optimal unbiased reconciliation. Second, the irrelevance of weight suggests that (9) reconciles the base forecast as if the forecast error variance of each variable can be minimized independently, but at the same time, (9) can be obtained by minimizing the variance of only the bottom-level variables. The extension suggests that these two apparently contradictory interpretations can coexist.

6. Conclusions

In this paper, we have provided an alternative proof to the minimum trace reconciliation developed by [1], filling a gap in their proof. We have also shown that an almost identical proof can be used to prove [5], so both the trace and weighted trace can be analyzed from a unified perspective. We believe the alternative simpler proof provides

additional insights and contributes to deepening the understanding of the minimum trace reconciliation.

References

1. Wickramasuriya, S.L.; Athanasopoulos, G.; Hyndman, R.J. Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization. *J. Am. Stat. Assoc.* **2019**, *114*, 804–819. [CrossRef]
2. Taieb, S.B.; Taylor, J.W.; Hyndman, R.J. Hierarchical Probabilistic Forecasting of Electricity Demand with Smart Meter Data. *J. Am. Stat. Assoc.* **2021**, *116*, 27–43. [CrossRef]
3. Ando, S.; Kim, T. Systematizing Macroframework Forecasting: High-Dimensional Conditional Forecasting with Accounting Identities. *IMF Econ. Rev.* **2023**. [CrossRef]
4. Athanasopoulos, G.; Gamakumara, P.; Panagiotelis, A.; Hyndman, R.J.; Affan, M. Hierarchical forecasting. In *Macroeconomic Forecasting in the Era of Big Data*; Springer: Berlin/Heidelberg, Germany, 2020.
5. Panagiotelis, A.; Athanasopoulos, G.; Gamakumara, P.; Hyndman, R.J. Forecast Reconciliation: A Geometric View with New Insights on Bias Correction. *Int. J. Forecast.* **2021**, *37*, 343–359. [CrossRef]
6. Horn, R.A.; Johnson, C.R. *Matrix Analysis*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2013.
7. Luenberger, D.G. *Optimization by Vector Space Methods*; John Wiley & Sons: Hoboken, NJ, USA, 1969.