INTERNATIONAL MONETARY FUND

An Alternative Proof of Minimum Trace Reconciliation

Sakai Ando and Futoshi Narita

WP/22/136

2022
JUL

WORKING PAPER

 WP/22/136

IMF Working Paper
Research Department

An Alternative Proof of Minimum Trace Reconciliation

Prepared by Sakai Ando and Futoshi Narita

Authorized for distribution by Prachi Mishra
July 2022

ABSTRACT: Minimum trace reconciliation, developed by Wickramasuriya et. al. (2019), is an innovation in the literature of forecast reconciliation. The proof, however, is indirect and not easy to extend to more general situations. This paper provides an alternative proof based on the first-order condition in the space of non-square matrix and argues 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.

RECOMMENDED CITATION: Ando, Sakai and Futoshi Narita (2022), “An Alternative Proof of Minimum Trace Reconciliation,” IMF Working Paper WP/22/136.

JEL Classification Numbers:	C53, E17
Keywords:	Forecast Reconciliation; Minimum Trace Reconciliation; Hierarchical Time Series
Author’s E-Mail Address:	sando@imf.org; fnarita@imf.org

WORKING PAPERS

An Alternative Proof of Minimum Trace Reconciliation

Prepared by Sakai Ando and Futoshi Narita

Contents

1. Introduction 3
2. Setup 3
3. Proof of (9) in Wickramasuriya et. al. (2019) 4
4. An Alternative Proof of (9) 5
5. An Extension of the Alternative Proof 6
6. Conclusion 8
References 9

1. Introduction

Minimum trace reconciliation, developed by Wickramasuriya et. al. (2019), 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 (Taieb et. al., 2021) and macroframework forecasting (Athanasopoulos et. al., 2020; Ando and Kim, 2022) to name a few. The proof of Wickramasuriya et. al. (2019), however, is indirect and not easy to extend to more general situations.

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, instead of finding a lower bound function and then solving a constrained minimization problem of the lower bound as in Wickramasuriya et. al. (2019). An almost identical proof can be used to prove Theorem 3.3 of Panagiotelis et. al. (2021), which shows that the minimum trace reconciliation and minimum weighted trace reconciliation lead to identical formula. By further extending the insights of the minimum weighted trace reconciliation, we can see why the lower bound minimization in Wickramasuriya et. al. (2019) 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 Wickramasuriya et. al. (2019). In section 4, we provide an alternative proof of Wickramasuriya et. al. (2019). Section 5 extends the proof to incorporate Panagiotelis et. al. (2021) and discusses the insights. In section 6, we conclude.

2. Setup

The setup and notation follow Wickramasuriya et. al. (2019). 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

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

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

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

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

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

$$\mathbb{E}_T \tilde{y}_T(h) = \mathbb{E}_T y_{T+h} = \mathbb{E}_T \hat{y}_T(h), \quad \forall y_{T+h} \Leftrightarrow SPS\mathbb{E}_T b_{T+h} = S\mathbb{E}_T b_{T+h}, \quad \forall \mathbb{E}_T b_{T+h} \Leftrightarrow SPS = S \Leftrightarrow PS = I_n. \tag{6}$$

Note that the necessity of the last equivalence follows from multiplying S' from left on 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 left on both sides.

The forecast error of the reconciled forecast can be expressed as

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

where $W = \mathbb{E}_T\left[\left(y_{T+h} - \hat{y}_T(h)\right)\left(y_{T+h} - \hat{y}_T(h)\right)'\right]$ 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

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

Wickramasuriya et. al. (2019) proves that the matrix P that minimizes the trace of the covariance matrix subject to the unbiasedness constraint is

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

3. Proof of (9) in Wickramasuriya et. al. (2019)

The proof of Wickramasuriya et. al. (2019) can be divided into two steps. First, the objective function is bounded from below using Weyl's inequalities

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

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

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

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

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.

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

$$\langle A, B\rangle = tr(A'B), \qquad A, B \in \mathbb{R}^{n\times m}. \tag{12}$$

By Theorem 1 of Luenberger (1969, p.243), there exists an $n \times m$ matrix Lagrange multiplier Λ such that the Lagrangian

$$L(P) = tr(SPWP'S') + tr\big(\Lambda'(I_n - PS)\big) \tag{13}$$

is stationary at its minimum point. This means that, at the minimum, the directional derivative (or Gateaux differential as defined on page 171 of Luenberger, 1969) of $L(P)$ is zero for any $n \times m$ matrix H.

$$\begin{aligned}\lim_{\alpha\to 0}\frac{L(P+\alpha H)-L(P)}{\alpha} &= tr(SHWP'S' + SPWH'S' - \Lambda'SHS)\\ &= tr(2SHWP'S' - \Lambda'SHS)\\ &= tr\big(H(2WP'S'S - S\Lambda'S)\big)\\ &= 0.\end{aligned} \tag{14}$$

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,

$$2WP'S'S = S\Lambda'S \Rightarrow P'S'S = \frac{1}{2}W^{-1}S\Lambda'S. \tag{15}$$

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

$$S'S = \frac{1}{2}S'W^{-1}S\Lambda'S \Rightarrow (S'W^{-1}S)^{-1}S'S = \frac{1}{2}\Lambda'S. \tag{16}$$

Thus, the formula follows.

$$P'S'S = W^{-1}S(S'W^{-1}S)^{-1}S'S \Rightarrow P = (S'W^{-1}S)^{-1}S'W^{-1}. \tag{17}$$

QED.

The proof essentially uses the extension of 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 Panagiotelis et. al. (2021). To motivate the extension, suppose we have the base forecast of the variables in the GDP expenditure approach that satisfy

$$\hat{Y} = \hat{C} + \hat{I} + \hat{G} + \widehat{XM}, \tag{18}$$

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

$$V\left(Y - \tilde{Y}\right) + V\left(C - \tilde{C}\right) + V(I - \tilde{I}) + V\left(G - \tilde{G}\right) + V\left(XM - \widetilde{XM}\right) \tag{19}$$

subject to the constraint (18). 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.

$$\omega_Y V\left(Y - \tilde{Y}\right) + \omega_C V\left(C - \tilde{C}\right) + \omega_I V(I - \tilde{I}) + \omega_G V\left(G - \tilde{G}\right) + \omega_{XM} V\left(XM - \widetilde{XM}\right), \qquad \omega_i > 0, \qquad \sum_i \omega_i = 1. \tag{20}$$

Such specification can be expressed as a weighted trace

$$\min_P tr(\omega SPWP'S') \ s.t.\ PS = I_n, \tag{21}$$

where ω is a $m \times m$ diagonal 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. As Panagiotelis et. al. (2021) show, the optimal matrix P is independent of ω as long as ω is symmetric and invertible.

Claim: For any symmetric and invertible $m \times m$ matrix ω, the solution to (21) is

$$P = (S'W^{-1}S)^{-1}S'W^{-1}. \tag{22}$$

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

$$L(P) = tr(\omega SPWP'S') + tr\big(\Lambda'(S - SPS)\big). \tag{23}$$

Since the Gateau derivative needs to be zero,

$$\begin{aligned}\lim_{\alpha\to 0}\frac{L(P+\alpha H)-L(P)}{\alpha} &= tr(\omega SHWP'S' + \omega SPWH'S' - \Lambda' SHS)\\ &= tr(2\omega SHWP'S' - \Lambda' SHS)\\ &= tr\big(H(2WP'S'\omega S - S\Lambda' S)\big)\\ &= 0.\end{aligned} \tag{24}$$

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

$$2WP'S'\omega S = S\Lambda' S \Rightarrow P'S'\omega S = \frac{1}{2}W^{-1}S\Lambda' S. \tag{25}$$

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

$$S'\omega S = \frac{1}{2}S'W^{-1}S\Lambda' S \Rightarrow (S'W^{-1}S)^{-1}S'\omega S = \frac{1}{2}\Lambda' S. \tag{26}$$

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

$$P'S'\omega S = W^{-1}S(S'W^{-1}S)^{-1}S'\omega S \Rightarrow P = (S'W^{-1}S)^{-1}S'W^{-1}. \tag{27}$$

QED.

Intuitively, the fact that the weight matrix doesn't matter can be interpreted as saying that there isn't a trade-off between variables. In other words, 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

$$tr(\omega SPWP'S') = tr(ASPWP'S'A') \tag{28}$$

for any full-rank square matrix A. In fact, since the proof only requires $S'\omega S$ to be invertible, one can further extend the claim and allow A to be a full-rank non-square matrix. For example, if $A = (S'S)^{-1}S'$, $S'A'AS = I_n$ is invertible, and the objective function collapses to (11)

$$tr(ASPWP'S'A') = tr(PWP'). \tag{29}$$

This is one way to see why the proof of Wickramasuriya et. al. (2019) reaches the same formula. One insight from the right side of (29) 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.

6. Conclusion

In this paper, we have provided an alternative proof to the minimum trace reconciliation developed by Wickramasuriya et. al. (2019). We have also shown that an almost identical proof can be used to prove Panagiotelis et. al. (2021), 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

Ando, Sakai, Taehoon Kim (2022), “Systematizing Macroframework Forecasting: High-Dimensional Conditional Forecasting with Accounting Identities,” *IMF Working Paper 22/110*.

Athanasopoulos, George, Puwasala Gamakumara, Anastasios Panagiotelis, Rob Hyndman, Mohamed Affan (2020), “Hierarchical forecasting,” *Macroeconomic Forecasting in the Era of Big Data.*

Luenberger, David G, (1969), “Optimization by Vector Space Methods,” *John Wiley & Sons*.

Panagiotelis, Anastasios, Puwasala Gamakumara, George Athanasopoulos, Rob Hyndman (2021), “Forecast Reconciliation: A Geometric View with New Insights on Bias Correction,” *International Journal of Forecasting.*

Taieb, Souhaib, James Taylor, Rob Hyndman (2021), “Hierarchical Probabilistic Forecasting of Electricity Demand with Smart Meter Data,” *Journal of American Statistical Association*.

Wickramasuriya, Shanika, George Athanasopoulos, Rob Hyndman (2019), “Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization,” *Journal of the American Statistical Association*.

PUBLICATIONS

An Alternative Proof of Minimum Trace Reconciliation
Working Paper No. **WP/22/136**