A Proofs

A.1 Lemma 1

Proof. Let the h-step-ahead base and reconciled forecast errors be given as

$$\hat{\boldsymbol{e}}_t(h) = \boldsymbol{y}_{t+h} - \hat{\boldsymbol{y}}_t(h)$$
$$\text{and} \qquad \tilde{\boldsymbol{e}}_t(h) = \boldsymbol{y}_{t+h} - \tilde{\boldsymbol{y}}_t(h),$$

for $t = 1, 2, \ldots$, where $\hat{\boldsymbol{y}}_t(h)$ and $\tilde{\boldsymbol{y}}_t(h)$ are the h-step-ahead base and reconciled forecasts using information up to and including time t, and $\boldsymbol{y}_{t+h}$ are the observed values of all series at time $t+h$.

We have that

$$\begin{aligned}
\tilde{\boldsymbol{e}}_t(h) &= \hat{\boldsymbol{e}}_t(h) + (\boldsymbol{I} - \boldsymbol{S}\boldsymbol{P})\,\hat{\boldsymbol{y}}_t(h) \\
&= \hat{\boldsymbol{e}}_t(h) + (\boldsymbol{I} - \boldsymbol{S}\boldsymbol{P})\,[\boldsymbol{y}_{t+h} - \hat{\boldsymbol{e}}_t(h)] \\
&= \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{e}}_t(h) + (\boldsymbol{I} - \boldsymbol{S}\boldsymbol{P})\boldsymbol{y}_{t+h}.
\end{aligned}$$

Since $\boldsymbol{y}_{t+h} = \boldsymbol{S}\boldsymbol{b}_{t+h}$, where $\boldsymbol{b}_{t+h}$ is the vector of observations at the bottom level, $(\boldsymbol{I} - \boldsymbol{S}\boldsymbol{P})\boldsymbol{S} = 0$. Hence $\tilde{\boldsymbol{e}}_t(h) = \boldsymbol{S}\boldsymbol{P}\hat{\boldsymbol{e}}_t(h)$ giving

$$\text{var}[\tilde{\boldsymbol{e}}_t(h)|\boldsymbol{\mathcal{I}}_t] = \boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}',$$

where $\boldsymbol{W}_h$ is the variance covariance matrix of the h-step-ahead base forecast errors. □

A.2 Theorem 1

Proof. Reformulate the objective function such that,

$$\text{tr}[\boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'\boldsymbol{S}'] = \text{tr}[\boldsymbol{S}'\boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'].$$

As $\boldsymbol{S}'\boldsymbol{S}$ and $\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'$ are both symmetric, and positive definite and positive semidefinite respectively, using Lemma 1 of Wang et al. (1986) we have

$$\text{tr}[\boldsymbol{S}'\boldsymbol{S}\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'] \geq \lambda_{\min}(\boldsymbol{S}'\boldsymbol{S})\text{tr}[\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'] \geq \lambda_{\min}(\boldsymbol{C}'\boldsymbol{C} + \boldsymbol{I})\text{tr}[\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'],$$

where $\lambda_{\min}(A)$ denotes the minimum eigenvalue of matrix $\boldsymbol{A}$ and $\boldsymbol{S} = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{I}_n \end{bmatrix}$. Since $\boldsymbol{C}'\boldsymbol{C}$ and $\boldsymbol{I}$ are both symmetric, and positive semidefinite and positive definite respectively, applying Weyl's inequalities (Horn & Johnson, 1990),

$$\lambda_{\min}(\boldsymbol{C}'\boldsymbol{C} + \boldsymbol{I})\mathrm{tr}[\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'] \geq [\lambda_{\min}(\boldsymbol{C}'\boldsymbol{C}) + \lambda_{\min}(\boldsymbol{I})]\mathrm{tr}[\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'] \geq \mathrm{tr}[\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'],$$

using the fact that all the eigenvalues of an identity matrix are unity.

Now $\boldsymbol{P}\boldsymbol{S} = \boldsymbol{I}_n$, so the minimization problem can be restated as

$$\min_{\boldsymbol{P}} \quad \mathrm{tr}[\boldsymbol{P}\boldsymbol{W}_h\boldsymbol{P}'] \qquad \text{such that} \quad \boldsymbol{P}\boldsymbol{S} = \boldsymbol{I}.$$

If $\boldsymbol{W}_h$ is positive definite and we let $\boldsymbol{L} = \boldsymbol{W}_h^{-\frac{1}{2}}\boldsymbol{S}$ and $\boldsymbol{H} = \boldsymbol{P}\boldsymbol{W}_h^{\frac{1}{2}}$, the minimization problem becomes

$$\min_{\boldsymbol{H}} \quad \mathrm{tr}[\boldsymbol{H}\boldsymbol{H}'] \qquad \text{such that} \quad \boldsymbol{H}\boldsymbol{L} = \boldsymbol{I}.$$

The unique solution of this well-known minimization problem is given by the Moore-Penrose generalized inverse of $\boldsymbol{L}$ (Penrose, 1956). As we know that $\boldsymbol{L}$ is full rank, $\boldsymbol{H} = (\boldsymbol{L}'\boldsymbol{L})^{-1}\boldsymbol{L}'$ which gives

$$\boldsymbol{P} = (\boldsymbol{S}'\boldsymbol{W}_h^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{W}_h^{-1}.$$

As the above unique solution of MinT has a similar representation to a GLS estimator of a least squares problem, we can reformulate the trace minimization problem in terms of a linear equality constrained least squares problem as follows:

$$\min_{\check{\boldsymbol{y}}_T(h)} \frac{1}{2}[\hat{\boldsymbol{y}}_T(h) - \check{\boldsymbol{y}}_T(h)]'\boldsymbol{W}_h^{-1}[\hat{\boldsymbol{y}}_T(h) - \check{\boldsymbol{y}}_T(h)]$$
$$\text{s.t.} \quad \check{\boldsymbol{y}}_T(h) = \boldsymbol{S}\check{\boldsymbol{b}}_T(h),$$

or equivalently,

$$\min_{\check{\boldsymbol{y}}_T(h)} \frac{1}{2}[\hat{\boldsymbol{y}}_T(h) - \check{\boldsymbol{y}}_T(h)]'\boldsymbol{W}_h^{-1}[\hat{\boldsymbol{y}}_T(h) - \check{\boldsymbol{y}}_T(h)]$$
$$\text{s.t.} \quad \boldsymbol{U}'\check{\boldsymbol{y}}_T(h) = \boldsymbol{0},$$

where $\check{\boldsymbol{b}}_T(h)$ is the vector of last n elements of $\check{\boldsymbol{y}}_T(h)$, $\boldsymbol{U}' = \begin{bmatrix} \boldsymbol{I}_{m^*} & -\boldsymbol{C} \end{bmatrix}$ and $m^* = m - n$.

The Lagrangian function for the minimization problem is

$$\mathcal{L}[\check{\boldsymbol{y}}_T(h), \boldsymbol{\lambda}] = \frac{1}{2}[\hat{\boldsymbol{y}}_T(h) - \check{\boldsymbol{y}}_T(h)]'\boldsymbol{W}_h^{-1}[\hat{\boldsymbol{y}}_T(h) - \check{\boldsymbol{y}}_T(h)] - \boldsymbol{\lambda}'\boldsymbol{U}'\check{\boldsymbol{y}}_T(h),$$

where $\boldsymbol{\lambda}$ is a Lagrange multiplier vector.

The first-order necessary conditions for $\tilde{\boldsymbol{y}}_T(h)$ to be a solution of the minimization problem require that there is a vector $\tilde{\boldsymbol{\lambda}}$ for which the following system of equations is satisfied:

$$\left[\begin{array}{c:c} \boldsymbol{W}_h^{-1} & \boldsymbol{U} \\ \hdashline \boldsymbol{U}' & \boldsymbol{0} \end{array}\right] \left[\begin{array}{c} \tilde{\boldsymbol{y}}_T(h) \\ \hdashline -\boldsymbol{\lambda} \end{array}\right] = \left[\begin{array}{c} \boldsymbol{W}_h^{-1}\hat{\boldsymbol{y}}_T(h) \\ \hdashline \boldsymbol{0} \end{array}\right].$$

Assuming that $\boldsymbol{W}_h^{-1}$ is positive definite yields the solution of the above system of equations as

$$\begin{aligned} \tilde{\boldsymbol{y}}_T(h) &= [\boldsymbol{I} - \boldsymbol{W}_h\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{W}_h\boldsymbol{U})^{-1}\boldsymbol{U}']\hat{\boldsymbol{y}}_T(h) \\ \text{and} \qquad \tilde{\boldsymbol{\lambda}} &= -(\boldsymbol{U}'\boldsymbol{W}_h\boldsymbol{U})^{-1}\boldsymbol{U}'\hat{\boldsymbol{y}}_T(h). \end{aligned}$$

The h-step-ahead reconciled forecasts at the bottom level are given by

$$\tilde{\boldsymbol{b}}_T(h) = \Big[\left[\,\boldsymbol{0}_{n\times m^*} \,\vdots\, \boldsymbol{I}_n\right] - \left[\,\boldsymbol{0}_{n\times m^*} \,\vdots\, \boldsymbol{I}_n\right] \boldsymbol{W}_h\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{W}_h\boldsymbol{U})^{-1}\boldsymbol{U}'\Big]\, \hat{\boldsymbol{y}}_T(h) := \boldsymbol{P}\hat{\boldsymbol{y}}_T(h).$$

Hence we can define an alternative representation for the $\boldsymbol{P}$ matrix by

$$\boldsymbol{P} = \boldsymbol{J} - \boldsymbol{J}\boldsymbol{W}_h\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{W}_h\boldsymbol{U})^{-1}\boldsymbol{U}',$$

where $\boldsymbol{J} = \left[\,\boldsymbol{0}_{n\times m^*} \,\vdots\, \boldsymbol{I}_n\right]$. □