Online supplement to "Optimal combination of forecasts with mean absolute error loss"

Felix Chan[1] and Laurent Pauwels[2]

[1]School of Economics and Finance, Curtin University, Perth, WA 6845, Australia. Email: Felix.Chan@cbs.curtin.edu.au.
[2]University of Sydney Business School, Darlington, NSW 2008, Australia and Centre for Applied Macroeconomic Analysis (CAMA), ANU. Email: laurent.pauwels@sydney.edu.au

August 2021

All numbers of equations, problems, theorems, corollaries, lemmas and propositions refer to those used in the paper "Optimal combination of forecasts with mean absolute error loss" by Felix Chan and Laurent Pauwels, unless otherwise specified.

Derivation of Problem (6)

The optimization problem of combining k forecasts at time t is expressed as

$$\begin{aligned} \text{minimize} \quad & \mathbb{E}\left|\nu_{0t} + \mathbf{u}_t^\top \mathbf{a}\right| \\ \text{subject to} \quad & \mathbf{1}^\top \mathbf{a} = 1, \end{aligned}$$

where the loss function is MAE and $\mathbb{E}$ denotes the expected value. The optimal solution to the above problem is the weight that minimizes the forecast specific idiosyncratic noises $\mathbf{u}_t = (u_{1t}, \ldots, u_{kt})^\top$. Let $\left(\mathbb{R}^{k+1}, \Im, G\right)$ be a probability space and define

$$\begin{aligned} X_{\mathbf{a}}^+ &= \left\{(\nu_{0t}, \mathbf{u}_t) : \nu_{0t} + \mathbf{u}_t^\top \mathbf{a} > 0\right\} \\ X_{\mathbf{a}}^- &= \left\{(\nu_{0t}, \mathbf{u}_t) : \nu_{0t} + \mathbf{u}_t^\top \mathbf{a} < 0\right\} \\ X_{\mathbf{a}}^0 &= \left\{(\nu_{0t}, \mathbf{u}_t) : \nu_{0t} + \mathbf{u}_t^\top \mathbf{a} = 0\right\}, \end{aligned}$$

then

$$\begin{aligned} &\mathbb{E}\left|\nu_{0t} + \mathbf{u}_t^\top \mathbf{a}\right| = \\ &\int_{X_{\mathbf{a}}^+} \left(\nu_{0t} + \mathbf{u}_t^\top \mathbf{a}\right) G\left(dv_t\right) - \int_{X_{\mathbf{a}}^-} \left(\nu_{0t} + \mathbf{u}_t^\top \mathbf{a}\right) G\left(dv_t\right) \\ &+ \int_{X_{\mathbf{a}}^0} \left(\nu_{0t} + \mathbf{u}_t^\top \mathbf{a}\right) G\left(dv_t\right), \end{aligned}$$

where $dv_t = d\nu_{0t} du_{1t} \ldots du_{kt}$. Note that the last integral is 0 because $\nu_{0t} + \mathbf{u}_t^\top \mathbf{a} = 0$ under the set $X_{\mathbf{a}}^0$. Moreover, under the assumption of stationarity

$$G\left(dv\right) = G\left(dv_t\right) \qquad \forall t.$$

As such, the subscript t is omitted from this point. The optimization problem above can then be restated as

$$\begin{aligned} \text{minimize} \quad & \int_{X_{\mathbf{a}}^+} \left(\nu_0 + \mathbf{u}^\top \mathbf{a}\right) G(dv) - \int_{X_{\mathbf{a}}^-} \left(\nu_0 + \mathbf{u}^\top \mathbf{a}\right) G(dv) \\ \text{subject to} \quad & \mathbf{1}^\top \mathbf{a} = 1. \end{aligned}$$

Derivation of Equation (10)

Equation (9) of Theorem 1 has a more intuitive representation. Define $\mathbb{1}_{-i}^+(\mathbf{a})$ as an indicator function such that $\mathbb{1}_{-i}^+(\mathbf{a}) = 1$ if $u_i < 0$ (indicated by the subscript) and $\mathbf{u} \in X_{\mathbf{a}}^+$ (indicated by the superscript) but 0 otherwise for all i. Similarly, $\mathbb{1}_{+i}^-(\mathbf{a})$ is an indicator function such that $\mathbb{1}_{+i}^-(\mathbf{a}) = 1$ if $u_i > 0$ and $\mathbf{u} \in X_{\mathbf{a}}^-$ but 0 otherwise. $\mathbb{1}_{-i}^-(\mathbf{a})$ and $\mathbb{1}_{+i}^+(\mathbf{a})$ are also defined in a similar manner. Then, by direct substitution the first order condition, $\boldsymbol{\omega}(\mathbf{a}) = 0$, can be written as:

$$\int u_i \left(\mathbb{1}_{+i}^+(\mathbf{a}^*) - \mathbb{1}_{+i}^-(\mathbf{a}^*)\right) G(dv) = \int u_i \left(\mathbb{1}_{-i}^+(\mathbf{a}^*) - \mathbb{1}_{-i}^-(\mathbf{a}^*)\right) G(dv).$$

This can be expressed more conveniently in terms of the expectation of $\mathbf{u}$ conditional on $\xi(a)$, namely,

$$\mathbb{E}\left[\mathbf{u}|\nu_0 + \mathbf{u}^\top \mathbf{a}^* > 0\right] \Pr\left(\nu_0 + \mathbf{u}^\top \mathbf{a}^* > 0\right) = \mathbb{E}\left[\mathbf{u}|\nu_0 + \mathbf{u}^\top \mathbf{a}^* < 0\right] \Pr\left(\nu_0 + \mathbf{u}^\top \mathbf{a}^* < 0\right).$$

Corollary to Theorem 1: Including the best forecast in combination set

The optimisation problem as stated in (6) intentionally excludes the best forecast, f_{0t}, from the combination set. This is because the inclusion of f_{0t} would lead to a corner solution as stated in Corollary 1.

Corollary 1. *If the best model, f_{0t}, is included in the choice set, then $\mathbf{a}^* = \mathbf{e}_1$.*

Corollary 1 implies that $\mathbb{E}\left|\nu_0\right| \leq \mathbb{E}\left|\nu_0 + \mathbf{u}^\top \mathbf{a}^*\right|$ since the best forecast would always minimize the MAE loss function. The result also shows that forecast combination will produce a forecast with lower mean absolute deviations than a single model since $\mathbf{a}^* \neq \mathbf{e}_i$ for any i in general.

Proof of Corollary 1: It is sufficient to verify that $\mathbf{e}_1$ satisfies equation (9). Rewrite $\nu_{0t} = \nu_{0t} + u_{0t}$ where $u_{0t} = 0$ for all t. Since v_{0t} and u_{it} are independent, it is obvious that

$\mathbb{E}(u_{it}|z_t > 0) = \mathbb{E}(u_{it}|z_t < 0) = 0$. Hence, $\mathbf{e}_i$ satisfies equation (9). This completes the proof. ■

Corollary to Theorem 1: Normal case

Corollary 2. *Denote* $\boldsymbol{\nu} = \left(\nu_0, \mathbf{u}^\top\right)^\top$, $\mathbb{E}\left(\mathbf{u}\mathbf{u}^\top\right) = \boldsymbol{\Omega}$ *and* $\boldsymbol{\nu} \sim N\left(\mathbf{0}, \boldsymbol{\Omega}_\nu\right)$ *with*

$$\boldsymbol{\Omega}_\nu = \begin{pmatrix} \sigma_\nu^2 & 0 \\ 0 & \boldsymbol{\Omega} \end{pmatrix}$$

then

$$\mathbf{a}^*_{\text{MAE}} = \mathbf{a}^*_{\text{MSE}}$$

where $\mathbf{a}^*_{\text{MAE}}$ *is the vector that satisfies equation (9) and* $\mathbf{a}^*_{\text{MSE}}$ *is the optimal weight vector when minimizing the Mean Square Errors (MSE) loss function. Specifically,* $\boldsymbol{\Omega}\mathbf{a}^*_{\text{MSE}} = \mathbf{1}\mathbf{a}^{*\top}_{\text{MSE}}\boldsymbol{\Omega}\mathbf{a}^*_{\text{MSE}}$ *as shown in Chan and Pauwels (2018).*

Proof of Corollary 2: Let $z = \nu_0 + \mathbf{u}^\top\mathbf{a}$ then $z \sim N(0, \sigma_z^2)$ with $\sigma_z^2 = \sigma_0^2 + \mathbf{a}^\top\Omega\mathbf{a}$. Moreover, let $\sigma_{iz} = \mathbb{E}(u_i z)$ then $\sigma_{iz} = \sigma_0 + \mathbf{e}_i\boldsymbol{\Omega}\mathbf{a}$ where $\mathbf{e}_i$ is a $k \times 1$ unit vector with the i^{th} element equals to 1 and 0 otherwise. Since the normal distribution is symmetrc around its mean, equation (9) implies that

$$\mathbb{E}(u_i|z > 0) - \mathbb{E}(u_i|z < 0) = \mathbb{E}(u_j|z > 0 - \mathbb{E}(u_j|z < 0).$$

Let $f_z(z|u_i)$ and $f_i(u)$ denote the conditional density of z conditional on u_i and the density of u_i, respectively, then

$$\begin{aligned} \mathbb{E}(u_i|z > 0) =& \frac{1}{2}\int_{-\infty}^{\infty}\int_0^{\infty} u f_z(z|u) f_i(u) dz du \\ =& \frac{1}{2\sqrt{2\pi\sigma_i^2}} \int_{-\infty}^{\infty} u \exp\left(-\frac{1}{2}\frac{u_i^2}{\sigma_i^2}\right) \Phi\left(\frac{\sigma_{iz}}{\left(\sigma_i^2\sigma_z^2 - \sigma_{iz}^2\right)^{\frac{1}{2}}}\frac{u_i}{\sigma_i}\right) du \\ =& \frac{\sigma_i}{2}\int_{-\infty}^{\infty} w\phi(w)\Phi\left(\delta_i w\right) dw \end{aligned} \tag{1}$$

where $\phi(x)$ and $\Phi(x)$ denote the normal density and the normal accumulative probability functions, respectively with

$$\delta_i = \frac{\sigma_{iz}}{\left(\sigma_i^2\sigma_z^2 - \sigma_{iz}^2\right)^{\frac{1}{2}}}.$$

Note that equation (1) is equivalent to $\mathbb{E}(w)$ if w follows a skew-normal distribution as defined in Azzalini (1985) with the skew parameter equals to δ_i, i.e. $w \sim SN(\delta_i)$. Therefore,

$$\begin{aligned}\mathbb{E}(u_i|z>0) =& \frac{\sigma_i^2}{2}\mathbb{E}(w)\\ =& \frac{1}{\sqrt{2\pi}}\frac{\sigma_{iz}}{\sigma_z}.\end{aligned}$$

Following the same arguments, it is straightforward to show that:

$$\mathbb{E}(u_i|z<0) = -\frac{1}{\sqrt{2\pi}}\frac{\sigma_{iz}}{\sigma_z}.$$

This implies

$$\mathbb{E}(u_i|z>0) - \mathbb{E}(u_i|z<0) = \frac{2}{\sqrt{2\pi}}\frac{\sigma_{iz}}{\sigma_z}.$$

Substitute this to equation (9) yields

$$\sigma_{iz} = \sigma_{jz}.$$

Since $\sigma_{iz}^2 = \sigma_0^2 + \mathbf{e}_i\mathbf{\Omega}\mathbf{a}^*$ and given the solutions to the optimisation problems for the MSE and MAE loss functions are unique, this implies:

$$\mathbf{\Omega}\mathbf{a}^* = \left(\mathbf{a}^{*\prime}\mathbf{\Omega}\mathbf{a}^*\right)\mathbf{1}. \tag{2}$$

This completes the proof. ■

1 Simulations: full results

1.1 Simulation framework

We generate five forecasts errors from skew Normal distributions only and then from multivariate t_3 distributions only. Two sets of optimal weights of the forecast combination obtained, one set by minimizing MAE loss and the other by minimizinng MSE loss. The simulations consist in comparing the optimal weights obtained from both minimization and provide evidence that the weights are equivalent as shown by the theory. The sample sizes are $N = \{20, 30, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000\}$. There are 5000 replications for each sample size. We now describe how we simulate random variates.[1]

[1]The simulation notebook with the code and results are available here.

1.1.1 Multivariate Skew Normal random variables

The multivariate skewed normal depends on three parameter matrices namely, a $p \times 1$ vector $\boldsymbol{\xi}$, a $p \times p$ matrix $\boldsymbol{\Lambda}$, and a $p \times p$ matrix $\boldsymbol{\Sigma}$. One can follow the steps below to simulate the p skewed normal random variates.

Step 1. Set the number of realisations (sample size) N we wish to simulate and the number of variates, p.

Step 2. Set the various parameter matrices namely, $\boldsymbol{\xi}$, $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}$.

Step 3. Simulate a $N \times p$ random matrix, $\mathbf{U}$, which follow the multivariate normal distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$.

Step 4. Simulate a $N \times p$ random matrix, $\boldsymbol{\tau}$, which follow the standard multivariate normal distribution $N(\mathbf{0}, \mathbf{I})$.

Step 5. Construct the multivariate skewed normal random vector as $\mathbf{Z}_{SN} = \boldsymbol{\xi}' \otimes \mathbf{i}_n + \boldsymbol{\Lambda}|\boldsymbol{\tau}| + \mathbf{U}$ where $|\boldsymbol{\tau}|$ denotes the absolute values of each element in $\boldsymbol{\tau}$.

The mean of $\mathbf{Z}_{SN}$ is

$$\mathbb{E}(\mathbf{Z}_{SN}) = \boldsymbol{\xi} + \sqrt{\frac{2}{\pi}}\boldsymbol{\Lambda}\mathbf{i} \tag{3}$$

where $\mathbf{i}$ is a vector of ones. This implies if we wish to have unbiased forecast, we should set

$$\boldsymbol{\xi} = -\sqrt{\frac{2}{\pi}}\boldsymbol{\Lambda}\mathbf{i}. \tag{4}$$

The variance-covariance matrix, $\boldsymbol{\Omega}$, of $\mathbf{Z}_{SN}$ is

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma} + \left(1 - \frac{2}{\pi}\right)\boldsymbol{\Lambda}\boldsymbol{\Lambda}' \tag{5}$$

For simplicity, we set $\boldsymbol{\Lambda}$ as a diagonal matrix which means $\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$ will also be a diagonal matrix with squares of the diagonal elements of $\boldsymbol{\Lambda}$. Note that these elements are not bounded as long as $\boldsymbol{\Omega}$ is a valid variance-covariance matrix i.e. positive semi-definite. We create a positive semi-definite matrix, $\boldsymbol{\Sigma}$ and a diagonal matrix $\boldsymbol{\Lambda}$. The elements in $\boldsymbol{\Lambda}$ can be either positive or negative, and even be 0 (which means that one would include some normal random variates in the mix). Now since $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$ are both positive definite, then the variance-covariance $\boldsymbol{\Omega}$ will also be positive semi-definite. The easiest way to generate a positive semi-definite matrix is to create a $p \times p$ matrix $\mathbf{A}$ and then set $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}'$. $\mathbf{A}$ can be created by drawing from a normal random generator. Here are the two sets of parameters we are using to produce the simulation results.

Set 1:

$$\mathbf{\Lambda} = \text{diag}\begin{pmatrix} 1.5 \\ -1 \\ 1.2 \\ -1.8 \\ 1.9 \end{pmatrix} \quad \text{and } \mathbf{\Sigma} = \begin{pmatrix} 6.16 & -1.51 & 3.34 & -4.41 & -3.88 \\ -1.51 & 10.49 & -3.31 & 2.47 & -0.99 \\ 3.34 & -3.31 & 7.77 & -1.10 & -2.96 \\ -4.41 & 2.47 & -1.10 & 11.13 & 4.00 \\ -3.88 & -0.99 & -2.96 & 4.00 & 8.61 \end{pmatrix}$$

which implies that the weights are $a_1 = 0.290$, $a_2 = 0.193$, $a_3 = 0.182$, $a_4 = 0.079$, and $a_5 = 0.255$.. The weights can be computed directly from the closed form solution $\mathbf{a}^* = \mathbf{\Omega}^{-1}\mathbf{i}\left(\mathbf{i}'\mathbf{\Omega}^{-1}\mathbf{i}\right)^{-1}$, where $\mathbf{\Omega} = \mathbf{\Sigma} + \left(1 - \frac{2}{\pi}\right)\mathbf{\Lambda}\mathbf{\Lambda}'$.

Set 2:

$$\mathbf{\Lambda} = \text{diag}\begin{pmatrix} -1 \\ -0.8 \\ 0.2 \\ -0.4 \\ 0.7 \end{pmatrix} \quad \text{and } \mathbf{\Sigma} = \begin{pmatrix} 6.11 & 3.18 & -0.99 & 1.41 & 2.16 \\ 3.18 & 7.63 & 3.70 & -2.82 & 1.32 \\ -0.99 & 3.70 & 3.51 & -3.40 & 0.19 \\ 1.41 & -2.82 & -3.40 & 4.59 & 0.78 \\ 2.16 & 1.32 & 0.19 & 0.78 & 3.40 \end{pmatrix}$$

which implies that the weights are $a_1 = 0.259$, $a_2 = -0.311$, $a_3 = 0.805$, $a_4 = 0.374$, and $a_5 = -0.128$.

1.1.2 Multivariate t random variables

The forecast errors generated with multivariate t_3 distributions also make use of the same $\mathbf{\Omega}$ variance-covariance matrices in both sets: $\mathbf{Z}_{t3} \sim t_3(\mathbf{0}, \mathbf{\Omega})$. Using the same $\mathbf{\Omega}$ variance-covariance matrices imply that the weights of the forecast combination with multivariate t_3 variates will be the same as the skew normal variates.

1.2 Simulation results

The extended results for Skew Normal forecast errors and t_3 forecast errors are presented here below. Only weights a_1 to a_4 are shown since $a_5 = 1 - \sum_{i=1}^{4} a_i$.

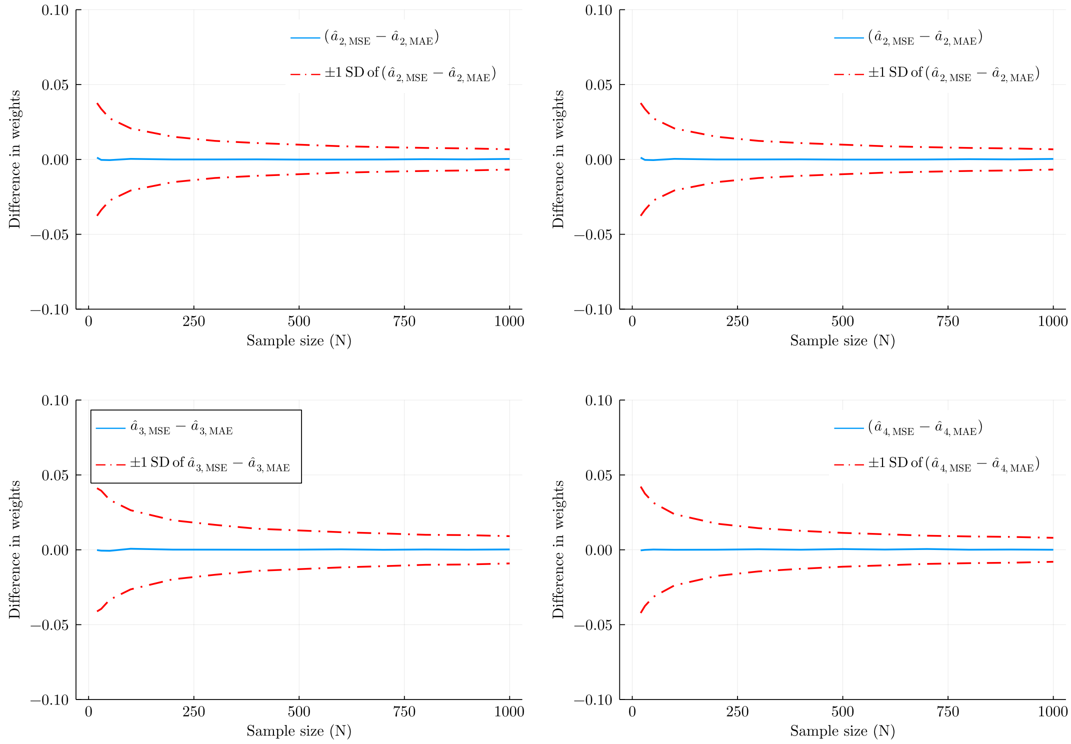

Figure 1: Skew Normal (Set 1)

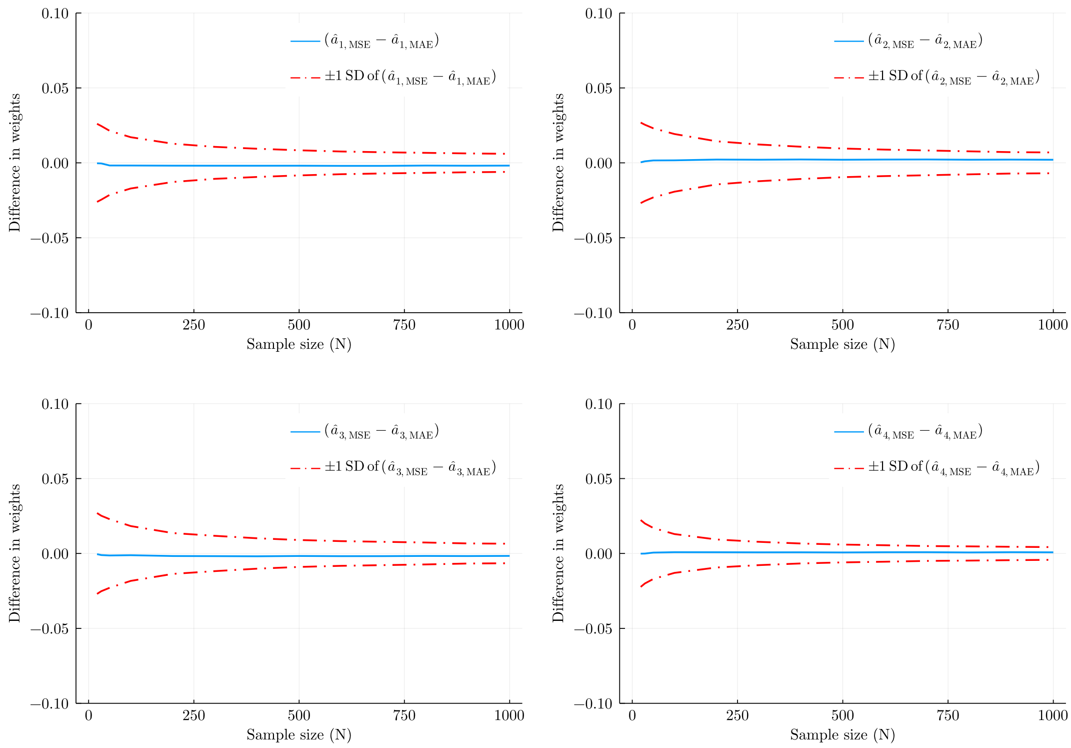

Figure 2: Skew Normal (Set 2)

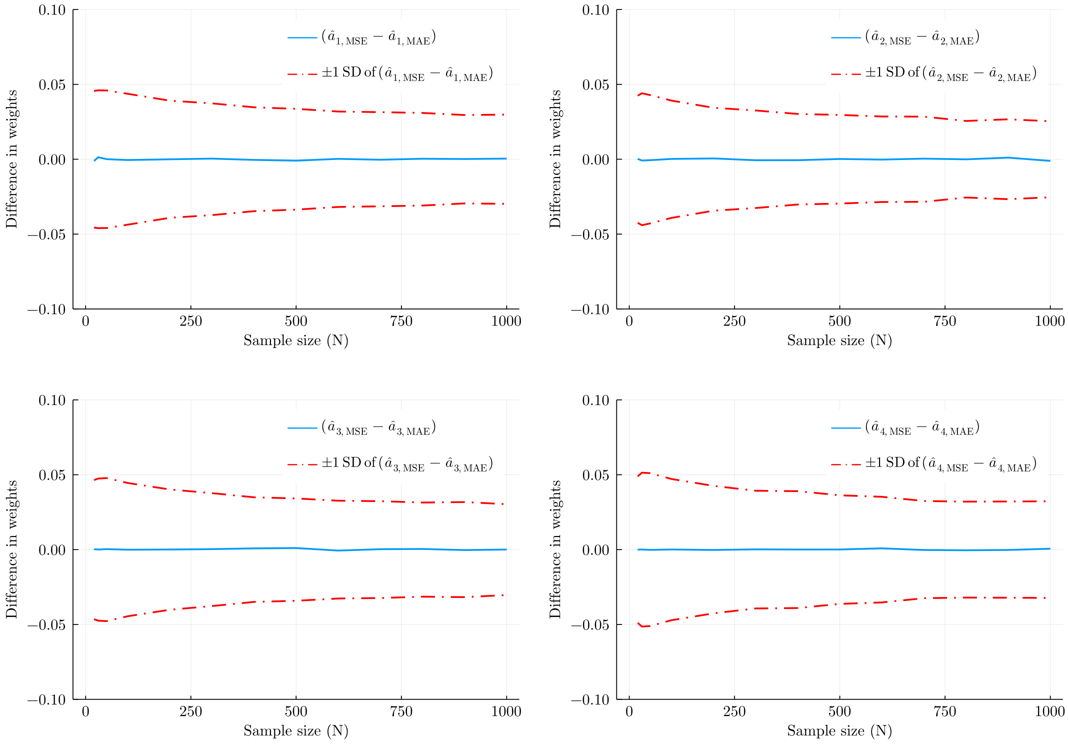

Figure 3: t_3 (Set 1)

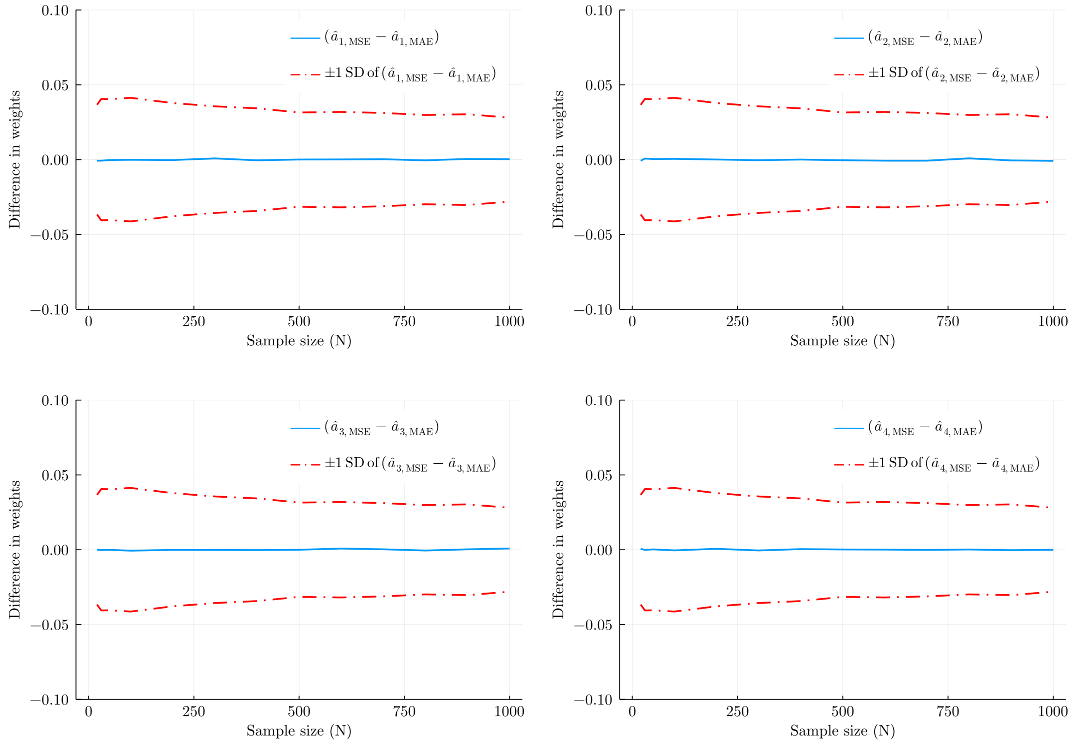

Figure 4: t_3 (Set 2)

REFERENCES

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, pages 171–178.

Chan, F. and Pauwels, L. L. (2018). Some theoretical results on forecast combinations. *International Journal of Forecasting*, 34(1):64–74.