COMBINATION OF FORECASTS

As the old saying goes, "Don't put all of your eggs in one basket lest you drop the basket and lose all of your eggs." Suppose the head of a forecasting division of a company has two sources of forecasts for the company's sales, one source being the forecasts generated by the division's econometrics group using an econometric time series model and the other source being the aggregated forecasts of the regional sales managers of the company. Suppose that the forecast horizon is h = 1 and represent the one-step-ahead forecasts of the econometrics group made at time t by $f_{t+1}^{(1)}$ and the forecasts of the managers as $f_{t+1}^{(2)}$. Now let us consider the statistical properties of the combination forecast

$$f_{t+1}^{(c)} = wf_{t+1}^{(1)} + (1-w)f_{t+1}^{(2)} \quad . \tag{1}$$

Suppose that both of the forecasting methods are **unbiased** in that the one-step-ahead forecast errors of the two methods, say $e_{t+1}^{(1)} = y_{t+1} - f_{t+1}^{(1)}$ and $e_{t+1}^{(2)} = y_{t+1} - f_{t+1}^{(2)}$, have zero expectations. That is, $E(e_{t+1}^{(1)}) = E(e_{t+1}^{(2)}) = 0$. Now consider the following theorem.

Theorem 1: Assume $f_{t+1}^{(1)}$ and $f_{t+1}^{(2)}$ are unbiased forecasts of y_{t+1}. Then the combination forecast $f_{t+1}^{(c)}$ is unbiased. That is, $E(e_{t+1}^{(c)}) = E(y_{t+1} - f_{t+1}^{(c)}) = 0$.

Proof: $E(e_{t+1}^{(c)}) = E(y_{t+1} - f_{t+1}^{(c)}) = E(y_{t+1} - wf_{t+1}^{(1)} - (1-w)f_{t+1}^{(2)})$

$$= E(w(y_{t+1} - f_{t+1}^{(1)}) + (1-w)(y_{t+1} - f_{t+1}^{(2)}))$$

$$= wE(y_{t+1} - f_{t+1}^{(1)}) + (1-w)E(y_{t+1} - f_{t+1}^{(2)})$$

$$= wE(e_{t+1}^{(1)}) + (1-w)E(e_{t+1}^{(2)}) = w \cdot 0 + (1-w) \cdot 0 = 0.$$ QED.

Another theorem follows straightforwardly.

Theorem 2: The variance of the combination forecast error is $Var(e_{t+1}^{(c)}) = w^2\sigma_{f1}^2 + (1-w)^2\sigma_{f2}^2 + 2w(1-w)\sigma_{f1f2}$, where $\sigma_{fi}^2 = E(e_{t+1}^{(i)})^2, i = 1,2$ are the variances of the forecast errors of the two methods and $\sigma_{f1f2} = E(e_{t+1}^{(1)}e_{t+1}^{(2)})$ is the covariance of the forecast errors of the two methods.

Proof: Note that $e_{t+1}^{(c)} = we_{t+1}^{(1)} + (1-w)e_{t+1}^{(2)}$. Therefore

$$E(e_{t+1}^{(c)})^2 = E[we_{t+1}^{(1)} + (1-w)e_{t+1}^{(2)}]^2$$

$$= w^2 E(e_{t+1}^{(1)})^2 + (1-w)^2 E(e_{t+1}^{(2)})^2 + 2w(1-w)E(e_{t+1}^{(1)} e_{t+1}^{(2)})$$

$$= w^2 \sigma_{f1}^2 + (1-w)^2 \sigma_{f2}^2 + 2w(1-w)\sigma_{f1f2} . \qquad \text{QED.}$$

As it turns out, there is a choice of the weight for the first forecasting method, say $w = w^*$, and thus for the weight of the second forecasting method, $(1-w^*)$, that minimizes the variance of the combination forecast error, $Var(e_{t+1}^{(c)})$. This result is stated in the following theorem.

Theorem 3: The choice of $w = w^* = \dfrac{\sigma_{f2}^2 - \sigma_{f1f2}}{\sigma_{f1}^2 + \sigma_{f2}^2 - 2\sigma_{f1f2}}$ minimizes the variance of the combination method forecast error.

Proof: The problem of solving for the optimal weight amounts to a first order condition problem in the calculus. Here the objective function is the variance of the combination forecast error, namely, $H(w) = E(y_{t+1} - f_{t+1}^{(c)})^2 = Var(e_{t+1}^{(c)})$. Thus, we want the minimize $H(w)$ as a function of w. That is,

$$\min_w H(w) = \min_w [w^2 \sigma_{f1}^2 + (1-w)^2 \sigma_{f2}^2 + 2w(1-w)\sigma_{f1f2}] \ .$$

The normal equation is

$$\frac{dH(w)}{dw} = \frac{d}{dw}[w^2 \sigma_{f1}^2 + (1-w)^2 \sigma_{f2}^2 + 2w(1-w)\sigma_{f1f2}]$$

$$= 2w\sigma_{f1}^2 - 2(1-w)\sigma_{f2}^2 + (2-4w)\sigma_{f1f2} \ .$$

The first order condition is

$$2w^* \sigma_{f1}^2 - 2(1-w^*)\sigma_{f2}^2 + (2-4w^*)\sigma_{f1f2} = 0.$$

Then, after some algebra, we get the optimal weight $w = w^* = \dfrac{\sigma_{f2}^2 - \sigma_{f1f2}}{\sigma_{f1}^2 + \sigma_{f2}^2 - 2\sigma_{f1f2}}$. The corresponding second order condition evaluated at $w = w^*$ is

$$\frac{d^2}{dw^2} H(w)_{|w=w^*} = 2\sigma_{f1}^2 + 2\sigma_{f2}^2 - 4\sigma_{f1f2} = 2(Var(e_{t+1}^{(1)} - e_{t+1}^{(2)})) > 0$$

which is positive for **all** values of w, much less at $w = w^*$. Thus, the second order condition for a minimum is met. QED.

Now for a discussion of the factors that affect the optimal weight, w^*, placed on the first method. First, we might note that the optimal weight doesn't require that the covariance between the errors of the forecasting methods (σ_{f1f2}) be negative as intuition might suggest. However, it can be easily shown that, for fixed values of the variances of the forecast errors of the two methods, σ_{f1}^2 and σ_{f2}^2, it is preferable to have a negative covariance rather than an positive covariance as the variance of the combination method's forecast error variance is smaller in the former case than in the latter case. Thus, given one unbiased forecasting method and the choice between one of two unbiased forecasting methods with equal forecasting on accuracy, i.e. forecast error variance, one would choose as a combining method the one with smaller positive covariance or, if possible, the one with the greatest (in an absolute value sense) negative covariance. In this way the benefits of forecast diversification are enhanced.

Second, note that if the two unbiased methods are equally accurate, $\sigma_{f1}^2 = \sigma_{f2}^2$, the optimal weights for the two unbiased forecasting methods are equal to 1/2 irregardless of the covariance between the errors of the two forecasting methods. Third, for a fixed accuracy of the first method and fixed covariance between the errors of the two methods, as the second forecasting method's accuracy declines, i.e. as $\sigma_{f2}^2 \to \infty$, the optimal weight w^* on the first method approaches one while the optimal weight on the second method, $(1 - w^*)$, approaches zero. This can be shown by applying L'Hospital's rule to the formula for w^* provided by Theorem 3. Of course, this argument is symmetric in the two methods. As $\sigma_{f1}^2 \to \infty$ and for fixed σ_{f2}^2 and σ_{f1f2}, the optimal weight on the second method $(1 - w^*)$ approaches one while the optimal weight on the first method approaches zero.

The next theorem presented here shows how diversification pays. When the optimal weights are chosen for the unbiased forecasting methods, the variance of the combination error is no greater than the smallest error variance between the two forecasting methods. Formally,

Theorem 4 (Diversification Pays): Let the optimal combination forecast made up of the two unbiased forecasts $f_{t+1}^{(1)}$ and $f_{t+1}^{(2)}$ be denoted

$$f_{t+1}^{(c^*)} = w^* f_{t+1}^{(1)} + (1 - w^*) f_{t+1}^{(2)}$$

with forecast error $e_{t+1}^{(c^*)} = y_{t+1} - f_{t+1}^{(c^*)}$. Then $Var(e_{t+1}^{(c^*)}) \leq Var(e_{t+1}^{(i)})$ for $i = 1,2$.

Proof: Classroom exercise.

ESTIMATION OF OPTIMAL COMBINATION WEIGHTS AND ADJUSTMENT FOR BIAS IN ONE OR MORE OF THE FORECASTS: THE NELSON AND GRANGER-RAMANATHAN COMBINATION METHODS

Nelson Method:

Assumption: Both methods produce unbiased forecasts. That is, $E\left(e_t^{(1)}\right) = E\left(e_t^{(2)}\right) = 0$.

The estimating equation for deriving the **Nelson combination (ensemble) weights** is

$$y_t = w f_t^{(1)} + (1 - w) f_t^{(2)} + \varepsilon_t \qquad (1)$$

or equivalently,

$$y_t - f_t^{(2)} = w(f_t^{(1)} - f_t^{(2)}) + \varepsilon_t. \qquad (2)$$

Now apply ordinary least squares to this transformed equation (2) and get $\widehat{w}$ for the estimate of w. Then the **Nelson Combination forecast (score)** is

$$\hat{y}_t^{(N)} = \widehat{w} f_t^{(1)} + (1 - \widehat{w}) f_t^{(2)}. \qquad (3)$$

Granger-Ramanathan Method:

Assumption: One or more methods produce biased forecasts. That is, for at least one of the methods, $E\left(e_t^{(i)}\right) \neq 0$.

The estimating equation for deriving the **Granger-Ramanathan combination (ensemble) weights** is

$$y_t = w_0 + w_1 f_t^{(1)} + w_2 f_t^{(2)} + \varepsilon_t \quad . \qquad (4)$$

Now apply ordinary least squares to equation (4) and get the coefficient estimates $\widehat{w}_0, \widehat{w}_1, \widehat{w}_2$ and form the combination (ensemble)

$$\hat{y}_t^{(GR)} = \widehat{w}_0 + \widehat{w}_1 f_t^{(1)} + \widehat{w}_2 f_t^{(2)}. \qquad (5)$$

This is called the **Granger-Ramanathan Combination forecast (score)**.