Journal of Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cjas20

On the Distribution of the Inverted Linear Compound of Dependent F-Variates and its Application to the Combination of Forecasts

Kuo-Yuan Liang [a] , Jack C. Lee [b] & Kurt S.H. Shao [c]
[a] Polaris Research Institute and Department of Economics , National Taiwan University , Taiwan
[b] National Chiao Tung University , Taiwan
[c] Polaris Research Institute , Taiwan
Published online: 01 Dec 2006.

Journal of Applied Statistics
Vol. 33, No. 9, 961–973, November 2006

On the Distribution of the Inverted Linear Compound of Dependent F-Variates and its Application to the Combination of Forecasts

KUO-YUAN LIANG*, JACK C. LEE** & KURT S.H. SHAO†

*Polaris Research Institute and Department of Economics, National Taiwan University, Taiwan, **National Chiao Tung University, Taiwan, †Polaris Research Institute, Taiwan*

ABSTRACT *This paper establishes a sampling theory for an inverted linear combination of two dependent F-variates. It is found that the random variable is approximately expressible in terms of a mixture of weighted beta distributions. Operational results, including rth-order raw moments and critical values of the density are subsequently obtained by using the Pearson Type I approximation technique. As a contribution to the probability theory, our findings extend Lee & Hu's (1996) recent investigation on the distribution of the linear compound of two independent F-variates. In terms of relevant applied works, our results refine Dickinson's (1973) inquiry on the distribution of the optimal combining weights estimates based on combining two independent rival forecasts, and provide a further advancement to the general case of combining three independent competing forecasts. Accordingly, our conclusions give a new perception of constructing the confidence intervals for the optimal combining weights estimates studied in the literature of the linear combination of forecasts.*

KEY WORDS: Combining weights, critical values, error-variance minimizing criterion, inverted F-variates, Pearson Type I approximation

Introduction

In this paper, we study the distribution of an inverted linear compound of dependent F-variates in the form:

$$\frac{1}{1 + a_1 F_1(T, T) + a_2 F_2(T, T)} \tag{1}$$

with degrees of freedom as indicated. Here, the two constants a_1 and a_2 lie in the interval (0,1]. This distribution is useful in constructing confidence intervals of the

Correspondence Address: Jack C. Lee, Graduate Institute of Finance, National Chio Tung University, Hsinchu, Taiwan. Email: jclee@stat.nctu.edu.tw.

0266-4763 Print/1360-0532 Online/06/090961–13 © 2006 Taylor & Francis
DOI: 10.1080/02664760600744330

minimum variance weights that can be attached to the components of the linear composite forecasts.

From Reid (1969), Dickinson (1973) or Newbold & Granger (1974), it is well known that given a history of unbiased forecast errors for k ($k \geq 2$) models, under the error-variance minimizing criterion, the optimal weighting vector (**W**) of the combined forecasts becomes

$$\mathbf{W} = \frac{\sum^{-1}\mathbf{u}}{\mathbf{u}'\sum^{-1}\mathbf{u}} \tag{2}$$

where **u** is a ($k \times 1$) vector of ones and $\sum$ a ($k \times k$) positive definite covariance matrix of forecasting errors between the k models.

It is worth noting that despite the popularity of equation (2), very little is known about the sampling properties of its estimator. A notable exception to this issue is the work of Dickinson (1973). When $k = 2$, based on the maximum likelihood estimator (**S**) of $\sum$ with zero off-diagonal elements and normally distributed forecasting errors, Dickinson (1973) demonstrated that each component of the estimated weight vector $\hat{\mathbf{W}}$ is expressible as a weighted beta or beta (if homoscedasticity is further imposed) distribution.

Although the combining procedure may involve more than two competing forecasts, we will restrict our attention to the $k = 3$ set-up with $\sum$ having zero off-diagonal elements and normally distributed forecasting errors only. This proves necessary as use of the general $k \times k$ set-up is technically difficult. Our restricted set-up hence extends the initial work of Dickinson (1973).

From the statistical viewpoint, a corollary of Dickinson's (1973) result is that an inverted F-variate of the form: $\frac{1}{1+aF(T,\,T)}$ where a is an arbitrary constant, is expressible as a weighted beta (if $a < 1$) or as a beta (if $a = 1$) distribution, i.e.,

$$\frac{1}{1 + aF(T,\,T)} \sim \text{weighted beta} \tag{3}$$

Another relevant theoretical contribution to our investigation is the work of Lee & Hu (1996). According to them, an arbitrary linear combination of two independent F-variates can be expressed approximately as a suitable constant (c) times an F density function, i.e.,

$$a_1F(u_1,\,u_2) + a_2F(u_1,\,u_2) \;\dot{\sim}\; cF(m_1,\,m_2) \tag{4}$$

where m_1 and m_2 are two positive constants.

Using the restricted set-up (detailed above), this paper derives the sampling distribution of the estimated combining weights. To achieve this goal, we begin in the next section with a reformulation of equation (2) by replacing $\sum$ with **S**, and show that each estimated weight is an inverted linear compound of dependent F-variates. We then relax the independence assumption on equation (4), and demonstrate in the third section that an expression of the right-hand side of equation (4) approximately still holds. Using this result and random variables transformation techniques, we also show in the third section that the distribution of an inverted linear compound of dependent F-variates of equation (1) is approximately expressible in terms of a mixture of weighted beta distributions. Additionally, we conduct extensive simulations to assess the accuracy of these approximations. Our results thus generalize those of Lee & Hu (1996) as well as Dickinson (1973). A notable implication of our theoretical results to the equal weighting scheme is elaborated as well.

Owing to the complexity of the derived distribution, the fourth section presents several operational results, including *r*th-order raw moments and critical values of the density based on the Pearson Type I approximation technique (Johnson *et al.*, 1963). The fifth section summarizes our findings and indicates future research directions.

Model and Related Results

Consider equation (2) in the restricted case where $\sum = diag(\sigma_{11}, \sigma_{22}, \sigma_{33})$. In practice, the parameters σ_{ii} are unknown, and $\sum$ is estimated by:

$$\mathbf{S} = diag\left(\frac{\sum_{t=1}^{T} e_{1t}^2}{T}, \frac{\sum_{t=1}^{T} e_{2t}^2}{T}, \frac{\sum_{t=1}^{T} e_{3t}^2}{T}\right)$$

where e_{it} is the error in the *t*th forecast value, using the *i*th forecasting method. Assuming e_{it} to be normally distributed with zero mean and variance σ_{ii}, the maximum likelihood estimator of $T\sum$ is given by:

$$T\mathbf{S} = diag\left(\sum_{t=1}^{T} e_{1t}^2, \sum_{t=1}^{T} e_{2t}^2, \sum_{t=1}^{T} e_{3t}^2\right). \tag{5}$$

It follows that:

$$\sum_{t=1}^{T} e_{it}^2 \sim \sigma_{ii}\chi^2(T) \tag{6}$$

From above, the *i*th weight in equation (2) is estimated by:

$$\hat{w}_i = \frac{1/\sum_{t=1}^{T} e_{it}^2}{\left(1/\sum_{t=1}^{T} e_{1t}^2\right) + \left(1/\sum_{t=1}^{T} e_{2t}^2\right) + \left(1/\sum_{t=1}^{T} e_{3t}^2\right)}, \quad i = 1, 2, 3 \tag{7}$$

Based on equation (7), each estimated weight can thus be written as an inverted linear compound of dependent *F*-variates of the form stated in equation (1). For example, equation (7) implies that another expression for the first estimated weight is:

$$\hat{w}_1 = \frac{1}{1 + \left(\sum_{t=1}^{T} e_{1t}^2 / \sum_{t=1}^{T} e_{2t}^2\right) + \left(\sum_{t=1}^{T} e_{1t}^2 / \sum_{t=1}^{T} e_{3t}^2\right)} \tag{8}$$

Using equation (6) in equation (8), we can conclude that:

$$\hat{w}_1 \sim \frac{1}{1 + a_1F_1(T, T) + a_2F_2(T, T)} \tag{9}$$

with degrees of freedom as indicated, and $a_1 = \sigma_{11}/\sigma_{22}$, $a_2 = \sigma_{11}/\sigma_{33}$ are two positive constants. Comparing this set-up with equation (1), we see that $\sigma_{11} \le \sigma_{22}$ and $\sigma_{11} \le \sigma_{33}$ are assumed here for illustrational convenience. Similar expressions for $\hat{w}_2$ and $\hat{w}_3$ can also be readily derived. Since $\sum_{t=1}^{T} e_{1t}^2$ appears in the second and third denominator terms of the right-hand expression of equation (8), it can be shown, that if $T > 4$ these two F-variates are indeed dependent and their correlation is given by:

$$corr(F_1, F_2) = \frac{T-4}{2(T-1)} \tag{10}$$

See Appendix A.1 for the proof.

Theorems, Simulations and Implications

Having verified the dependency between F_1 and F_2 in equation (10), we now turn our attention to the problem of finding the probability density of its linear compound of $a_1F_1(T, T)$ and $a_2F_2(T, T)$. As the following result indicates, this linear compound can be approximated by a constant (η) times an $F(m_1, m_2)$ variate, with the degrees of freedom as indicated.

Theorem 1

$$a_1F_1(T, T) + a_2F_2(T, T) \;\dot{\sim}\; \eta F(m_1, m_2) \tag{11}$$

where the right-hand expression of equation (11) comes from the denominator of equation (9), and the parameters η, m_1 and m_2 can be expressed in explicit forms in terms of a_1, a_2 and T. Specifically (Lee & Hu, 1996),

$$\eta = \frac{2A^2C - 2AB^2}{A^2B + 3AC - 4B^2}$$

$$m_1 = \frac{4A^2C - 4AB^2}{AB^2 - 2A^2C + BC}$$

and

$$m_2 = \frac{2A^2B + 6AC - 8B^2}{A^2B + AC - 2B^2} \tag{12}$$

where

$$A = \frac{(a_1 + a_2)T}{T-2}$$

$$B = \frac{T(T+2)}{T-2}\left(\frac{a_1^2 + a_2^2}{T-4} + \frac{2a_1a_2}{T-2}\right)$$

$$C = \frac{T(T+2)(T+4)}{(T-2)(T-4)}\left(\frac{a_1^3 + a_2^3}{T-6} + \frac{3a_1^2a_2 + 3a_1a_2^2}{T-2}\right) \tag{13}$$

and $T > 6$.

Similar to Lee & Hu (1996), we conduct an extensive simulation study to assess the accuracy of this approximation. The results of our study are summarized in Table 1.

In the simulation, we conduct 15,000 runs for each linear compound of the form $a_1F_1(T, T) + a_2F_2(T, T)$ and compute the probabilities of exceeding the 1%, 5% and 10% points. From Table 1, we see that the approximation to the assigned probability

Table 1. Simulated results for Theorem 1

	Tail probability		
Linear compound	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
$F_1(10,10) + F_2(10,10)$	0.0101	0.0510	0.1019
$F_1(10,10) + 0.5F_2(10,10)$	0.0101	0.0510	0.1024
$0.3F_1(10,10) + 0.01F_2(10,10)$	0.0106	0.0491	0.0994
$0.7F_1(10,10) + 0.01F_2(10,10)$	0.0103	0.0493	0.0993
$0.9F_1(10,10) + 0.01F_2(10,10)$	0.0097	0.0491	0.1000
$F_1(10,10) + 0.01F_2(10,10)$	0.0098	0.0500	0.0984
$0.9F_1(10,10) + 0.7F_2(10,10)$	0.0106	0.0507	0.1039
$0.9F_1(30,30) + 0.01F_2(30,30)$	0.0098	0.0493	0.0978
$F_1(30,30) + 0.01F_2(30,30)$	0.0098	0.0493	0.0972

Table entries are the simulated probabilities in the right-hand tail of the listed linear compound of dependent F-variates.

(α) in the right-hand tail of the listed linear compound of dependent F-variates is generally quite accurate.

By virtue of Theorem 1, $\hat{w}_1$ in equation (9) can thus be reasonably approximated by:

$$\hat{w}_1 \ \dot{\sim} \ \frac{1}{1 + \eta F(m_1, m_2)} \tag{14}$$

Likewise, similar expressions for $\hat{w}_2$ and $\hat{w}_3$ can be obtained. Using the approximations derived in equation (14), we are ready to apply the variable transformation techniques to derive the probability density of $f(\hat{w}_i)$ $(i = 1, 2, 3)$.

Theorem 2

Let $e_{it}(i = 1, 2, 3; t = 1, 2, \ldots, T)$ be the error in the tth forecast value using the ith forecasting model. Assume at a particular point of time, $\mathbf{e}' = (e_{1t}, e_{2t}, e_{3t}) \sim N(\mathbf{0}, \sum)$ with $\sum = diag(\sigma_{11}, \sigma_{22}, \sigma_{33})$, then under the error-variance minimizing criterion, the distribution of the ith optimal combining weight estimator $\hat{w}_i$ is approximately a mixture of beta random variables with the probability density function of the form:

$$f(\hat{w}_i) = \frac{(1-b)^{m_2/2}}{B(m_2/2, m_1/2)} \sum_{j=0}^{\infty} \left[C_j^{m+j-1} b^j B\left(\frac{m_2}{2} + j, \frac{m_1}{2}\right) Beta\left(\frac{m_2}{2} + j, \frac{m_1}{2}\right) \right] \tag{15}$$

where $C_j^{m+j-1} = \frac{(m+j-1)!}{j!(m-1)!}$, $b = 1 - \frac{\eta m_2}{m_1}$, $B(p, q)$ is a beta function and $Beta(p, q)$ is a beta density function with parameters p, q, respectively.

Proof

For the proof see Appendix A.2.

The following two theorems show that the condition $|b| < 1$ is sufficient for the integrability of $f(\hat{w}_i)$ over (0,1], the satisfaction of $\int_0^1 f(\hat{w}_i) d\hat{w}_i = 1$, and the existence of the rth order raw moment.

Theorem 3

If $|b| < 1$, then $\int_0^1 f(\hat{w}_i)d\hat{w}_i = 1$.

Proof

For the proof see Appendix A.3.

Theorem 4

If $|b| < 1$, then the rth-order raw moment of $f(\hat{w}_i)$ exists.

Proof

For the proof see Appendix A.4.

A notable implication of the condition $|b| < 1$ is elaborated as follows. Suppose the matrix $\sum$ in equation (2) is further restricted to $\sum = diag(\sigma_{11}, \sigma_{22}, \sigma_{33})$ and $\sigma_{11} = \sigma_{22} = \sigma_{33}$. It would immediately appear that this case would generate $w_1 = w_2 = w_3 = 1/3$. The case in which each forecast receives equal weight is of particular interest, because it may be reasonable in many realistic applications.

More specifically, the usual rationale for the equal weighting scheme is as follows. First, 'if (a) there is only a small data base and/or (b) the error covariance structure is not stationary' (Bunn, 1986, p. 152), then specifying $\sum$ as an unrestricted real symmetric positive definite matrix tends to cause the robustness problems due to poor estimation of its elements. A resolution is therefore suggested to specify the matrix $\sum$ in our restricted setup as $diag(\sigma_{11}, \sigma_{22}, \sigma_{33})$. Second, if no information is known or no reason to believe a priori on the relative accuracy of the competing forecasts, an even more extreme response is to further impose the constraint $\sigma_{11} = \sigma_{22} = \sigma_{33}$ into the above diagonal setting and utilize the equal weighting scheme (Bunn, 1986).

This extreme case means that $a_1 = a_2 = 1$ in equation (9). Substituting $a_1 = a_2 = 1$ for b in equation (15) and using the condition $|b| < 1$ produces:

$$0 < \frac{4(T-6)(3T^2 - 10T - 10)}{(T-2)(3T^2 - 16T + 28)} < 2$$

Significantly, the above inequality holds only when $T = 7,8,9$. Therefore, as a practical matter, the existing conditions of $f(\hat{w}_i)$ and its rth-order moment in this particular equal weighting scheme are extremely hard to satisfy. A cautious approach is suggested when applying this method, where other sources also share this view (Bunn, 1986; Winkler & Clemen, 1992)

Theorem 4 gives the following corollary.

Corollary 1

Each rth-order moment of $f(\hat{w}_i)$ is expressible as a monotonically decreasing sequence.

Proof

For the proof see Appendix A.5

To check the validity of the properties expressed in Theorems 2, 3, 4 and Corollary 1, the raw moments of $\hat{w}_i$ up to the fourth-order with sample sizes 10, 30 and 100 are studied separately. Because each of the three weights studied leads to similar conclusions, only the numerical results of the first weight (w_1) are reported in Table 2. Three important points are noted as follows. First, for each reported inverted linear compound with three different sample sizes, the condition $|b| < 1$ is satisfied. For example, for the case, $1/(1 + 0.7 F_1 + 0.01 F_2)$ with sample sizes 10, 30, and 100, $|b| = 0.5804, 0.4999$, and 0.4971, respectively. Second, the numerical results are consistent with Corollary 1, displaying a monotonically decreasing sequence pattern. Third, all $E(\hat{w}_1)$ entries have a downward bias, i.e. $E(\hat{w}_1) = E(1/1 + a_1F_1 + a_2F_2) < (1/1 + a_1 + a_2) = w_1$. However, the magnitude of this downward bias shrinks as the sample size increases, implying that $\hat{w}_1$ tends to be asymptotically unbiased for w_1.

Pearson Type I Approximation

Although the density of $\hat{w}_i$ has been derived in the previous section, the critical values for interval estimation and hypothesis testing purposes are still extremely hard to obtain. However, since the moments are available, we can approximate the distribution of $\hat{w}_i$ by the Pearson type I distribution which is defined as (Lee & Hu, 1996)

$$f(x) = [\beta(a+1, b+1)(\sigma_1 - \sigma_0)^{a+b+1}]^{-1}(x - \sigma_0)^a(\sigma_1 - x)^b$$

where $\sigma_0 \leq x \leq \sigma_1$, $a, b \in \mathbf{R}$.

Table 2. w_1 and the raw moments of $\hat{w}_1$ up to the fourth-order

$\hat{w}_1$	w_1	$E\hat{w}_1$	$E\hat{w}_1^2$	$E\hat{w}_1^3$	$E\hat{w}_1^4$
			Sample size 10		
$1/(1+0.02F_1+0.01F_2)$	0.970874	0.964700	0.931100	0.899077	0.868543
$1/(1+0.002F_1+0.001F_2)$	0.997009	0.996307	0.992634	0.988980	0.985345
$1/(1+0.3F_1+0.01F_2)$	0.763359	0.744231	0.567644	0.441612	0.349256
$1/(1+0.7F_1+0.01F_2)$	0.584795	0.576654	0.354011	0.228156	0.152900
$1/(1+0.9F_1+0.01F_2)$	0.523561	0.520868	0.293674	0.156437	0.095475
$1/(1+F_1+0.01F_2)$	0.497512	0.497211	0.269679	0.156437	0.095745
$1/(1+0.9F_1+0.7F_2)$	0.384618	0.382487	0.162398	0.075405	0.037481
$1/(1+F_1+0.5F_2)$	0.400000	0.397275	0.174562	0.083274	0.042524
			Sample size 30		
$1/(1+0.02F_1+0.01F_2)$	0.970874	0.968968	0.939000	0.910054	0.882093
$1/(1+0.002F_1+0.001F_2)$	0.997009	0.996798	0.993607	0.990427	0.987259
$1/(1+0.3F_1+0.01F_2)$	0.763359	0.756860	0.577213	0.443337	0.342773
$1/(1+0.7F_1+0.01F_2)$	0.584795	0.584420	0.346758	0.210638	0.130330
$1/(1+0.9F_1+0.01F_2)$	0.523561	0.522624	0.281083	0.155174	0.087736
$1/(1+F_1+0.5F_2)$	0.497512	0.497424	0.255416	0.134998	0.073263
			Sample size 100		
$1/(1+0.02F_1+0.01F_2)$	0.970874	0.970324	0.941556	0.913665	0.886626
$1/(1+0.3F_1+0.01F_2)$	0.763359	0.761404	0.581016	0.444323	0.340508
$1/(1+0.7F_1+0.01F_2)$	0.584795	0.583922	0.343273	0.203130	0.120970
$1/(1+0.9F_1+0.01F_2)$	0.523561	0.523275	0.276255	0.147106	0.078992
$1/(1+F_1+0.5F_2)$	0.497512	0.497487	0.249945	0.126783	0.064911

Table 3. Critical values for $\hat{w}_1$ based on the Pearson Type I approximation

		Tail Probability α							
$\hat{w}_1$	w_1	$\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.025$	$\alpha = 0.05$	$\alpha = 0.95$	$\alpha = 0.975$	$\alpha = 0.99$	$\alpha = 0.995$
		Sample size 10							
$1/(1 + 0.3F_1 + 0.01F_2)$	0.763359	0.361607	0.406223	0.470608	0.524328	0.902542	0.918216	0.932646	0.940401
$1/(1 + 0.7F_1 + 0.01F_2)$	0.584795	0.185902	0.223393	0.279248	0.327863	0.810466	0.849467	0.892380	0.920052
$1/(1 + F_1 + 0.5F_2)$	0.497512	0.125202	0.142536	0.171396	0.199804	0.625318	0.668873	0.717642	0.749386
		Sample size 30							
$1/(1 + 0.02F_1 + 0.01F_2)$	0.970874	0.934439	0.939205	0.945599	0.950575	0.982675	0.984428	0.986272	0.987420
$1/(1 + 0.3F_1 + 0.01F_2)$	0.763359	0.584444	0.598797	0.614514	0.642355	0.860101	0.874330	0.888617	0.896978
$1/(1 + 0.7F_1 + 0.01F_2)$	0.584795	0.348087	0.371800	0.405815	0.435148	0.721016	0.744503	0.770518	0.787430
		Sample size 100							
$1/(1 + 0.02F_1 + 0.01F_2)$	0.970874	0.954803	0.956636	0.959200	0.961292	0.978035	0.979237	0.980566	0.981427
$1/(1 + 0.3F_1 + 0.01F_2)$	0.763359	0.651932	0.664615	0.682600	0.697790	0.814163	0.821643	0.829568	0.834512

Table entries are the critical values with the probability α lying beneath.

In order to utilize the Pearson Type I approximation, we need the first four moments of $\hat{w}_i$, which can be obtained as demonstrated numerically in Table 2. Let $\mu = E(\hat{w}_i)$, $\mu_h = E(\hat{w}_i - \mu)^h$, $h = 2, 3, 4$ and $\beta_1 = \mu_3^2/\mu_2^3$, $\beta_2 = \mu_4/\mu_2^2$. Then the Pearson Type I distribution requires that $6 + 3\beta_1 - 2\beta_2 > 0$, $\beta_2 - \beta_1 - 1 > 0$.

Instead of computing from the density directly, we will make use of the tables produced by Johnson *et al.* (1963). For this purpose, we need the following double entry interpolations. Linear interpolation is often possible for β_2, while second differences are needed for $\sqrt{\beta_1}$. This procedure allows us to interpolate first for β_2 at each of the nearest four values of $\sqrt{\beta_1}$. Furthermore, it also tabulates first x_{-1}, x_0, x_1, x_2, and finally to interpolate for $\sqrt{\beta_1}$, using the formula

$$x(\theta) = (1 - \theta)x_0 + \theta x_1 - \frac{1}{4}\theta(1 - \theta)[\Delta^2 x_0 + \Delta^2 x_1]$$

where θ is the appropriate fraction in the tabular interval.

Based on the Pearson Type I approximation, as briefed above, Table 3 gives critical values of $\hat{w}_1$ with $\alpha = 0.005, 0.01, 0.025, 0.05, 0.95, 0.975, 0.99, 0.995$ for several cases considered in Table 2. In reference to Table 3, two major results emerged. First, by picking up $\alpha = 0.025$ and $\alpha = 0.975$, it can be seen that with sample sizes 10, 30 and 100 the 95% interval estimates of w_1 for $\hat{w}_1 = 1/(1 + 0.3F_1 + 0.01F_2)$ lie in the interval [0.470608, 0.918216], [0.614514, 0.874330] and [0.6826, 0.821643], respectively. Most importantly by using the data in Table 3, the same method also applies to the construction of interval estimates of w_1 based on particular $\hat{w}_1$ and distinctive width considerations. Second, as expected, we note that, under the preassigned percentage, the larger are the sample sizes the narrower are the interval weight estimates.

Conclusions

Among methods of combining forecasts (Liang, 1992), the formula (2) proposed by Reid (1969), Dickinson (1973) or Newbold & Granger (1974) is perhaps the single most extensively used measure of the optimal weights. Despite the popularity of this formula, very little is known about the sampling properties of its estimator. Although Dickinson (1973) has studied this issue, it only dealt with the combination of two forecasts exhibiting no covariance between their errors. Dickinson (1973, p. 259) also mentioned that '[t]he exact derivation of confidence intervals for the weights ... of the combined forecasts is extremely complex when more forecasts, or covariance between errors, are introduced'.

In this paper, attention has been directed mainly to the combination of three forecasts exhibiting no covariance between their errors. With normally distributed forecasting errors, we show that each estimated weight is expressible as an inverted linear compound of dependent *F*-variates and has approximately a mixture of weighted beta distributions.

Operational results, including *r*th-order raw moments and critical values of the density are subsequently obtained by using the Pearson Type I approximation technique. As a contribution to the probability theory, our findings extend Lee & Hu's (1996) recent investigation on the distribution of the linear compound of two independent *F*-variates. In terms of relevant applied works, our results refine Dickinson's (1973) inquiry on the distribution of the optimal combining weights estimates based on combining two independent rival forecasts, and provide a further advancement to the general case of combining three independent competing forecasts. Accordingly, this paper gives a new perception of constructing the confidence intervals for the optimal combining weights estimates studied in the literature of the linear combination of forecasts. A cautious approach is also suggested when applying the popular equal weighting combining method, because the existing conditions of $f(\hat{w}_i)$ and its rth-order moment are practically very hard to satisfy.

In this paper, we have enlarged the forecasting error covariance matrix from Dickinson's (1973) 2×2 diagonal setting to 3×3. We strongly hope that this can serve as a stepping stone into the studies based on the more general formulation of the matrix.

Appendix: Proofs

Appendix A.1. Proof of equation (10)

Suppose X_1, X_2 and X_3 are three independent chi-square distributed random variables with T degrees of freedom. By independence, we have

$$f_{X_1X_2X_3}(x_1,x_2,x_3) = f_{X_1}(x_1)f_{X_2}(x_2)f_{X_3}(x_3)$$

Using the following transformations of variables

$$F_1 = \frac{X_1}{X_2},\ F_2 = \frac{X_1}{X_3},\ F_3 = X_1$$

we obtain the joint probability density function of F_1, F_2 and F_3

$$f_{F_1F_2F_3}(f_1, f_2, f_3) = \frac{1}{[\Gamma(T/2)]^3 2^{(3T/2)}} f_1^{-(T/2)-1} f_2^{-(T/2)-1} f_3^{(3T/2)-1} e^{-(1+(1/f_1)+(1/f_2))f_3/2)},$$

and the joint probability density function of F_1 and F_2

$$f_{F_1F_2} = \frac{\Gamma(3T/2)}{[\Gamma(T/2)]^3} \frac{f_1^{-(T/2)-1} f_2^{-(T/2)-1}}{(1 + (1/f_1) + (1/f_2))^{(3T/2)}}$$

It is easy to verify that

$$E(F_j^{r_j}) = \frac{\Gamma(T/2 + r_j)\Gamma(T/2 - r_j)}{[\Gamma(T/2)]^2},\ \forall r_j \in \mathbf{N}$$

and

$$E(F_1^{r_1} F_2^{r_2}) = \frac{\Gamma(T/2 + r)\Gamma(T/2 - r_1)\Gamma(T/2 - r_2)}{[\Gamma(T/2)]^3},\ r = r_1 + r_2$$

Hence,

$$E(F_j) = \frac{T}{T-2}$$

$$E(F_j^2) = \frac{T(T+2)}{(T-2)(T-4)}$$

$$E(F_1F_2) = \frac{T(T+2)}{(T-2)^2}$$

and

$$\mathrm{cov}(F_1, F_2) = \frac{2T}{(T-2)^2}$$

$$\mathrm{var}(F_j) = \frac{4T(T-1)}{(T-2)^2(T-4)}$$

$$\mathrm{corr}(F_1, F_2) = \frac{T-4}{2(T-1)}$$

Appendix A.2. Proof of Theorem 2

Since

$$\hat{w}_i \sim \frac{1}{1 + a_1F_1 + a_2F_2} \dot{\sim} \frac{1}{1 + \eta F(m_1, m_2)}$$

we have the following approximate probability density function of $\hat{w}_i$

$$f(\hat{w}_i) = \frac{(1-b)^{m_2/2}}{B(m_2/2, m_1/2)} \frac{\hat{w}_i^{(m_2/2)-1}(1-\hat{w}_i)^{(m_1/2)-1}}{(1-b\hat{w}_i)^m}$$

Using a negative binomial expansion we can express $f(\hat{w}_i)$ as

$$f(\hat{w}_i) = \frac{(1-b)^{m_2/2}}{B(m_2/2, m_1/2)} \sum_{j=0}^{\infty} C_j^{(m+j)-1} b^j \hat{w}_i^{(m_2/2)+(j-1)} (1-\hat{w}_i)^{(m_1/2)-1}$$

or alternatively as

$$f(\hat{w}_i) = \frac{(1-b)^{m_2/2}}{B(m_2/2, m_1/2)} \sum_{j=0}^{\infty} \left[C_j^{(m+j)-1} b^j B\left(\frac{m_2}{2} + j, \frac{m_1}{2}\right) Beta\left(\frac{m_2}{2} + j, \frac{m_1}{2}\right)\right]$$

Appendix A.3. Proof of Theorem 3

$$\int_0^1 f(\hat{w}_i)\mathrm{d}\hat{w}_i = \int_0^1 \left[\frac{(1-b)^{(m_2/2)}}{B(m_2/2, m_1/2)} \sum_{j=0}^{\infty} C_j^{m+j-1} b^j B\left(\frac{m_2}{2} + j, \frac{m_1}{2}\right) Beta\left(\frac{m_2}{2} + j, \frac{m_1}{2}\right)\right] \mathrm{d}\hat{w}_i$$

$$= (1-b)^{m_2/2} \sum_{j=0}^{\infty} \left[C_j^{(m+j)-1} b^j \frac{B((m_2/2)+j,\, m_1/2)}{B(m_2/2,\, m_1/2)} \right]$$

Consider a non-negative infinite series $\{a_n\}$, and let

$$R = \limsup_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right|$$

If $R < 1$, then $\{a_n\}$ is absolutely convergent, by the ratio test (Apostol, 1974, p.173). By the previous theorem, with

$$R = \limsup_{n\to\infty} \left| \frac{m_2 + 2n}{2 + 2n} b \right| = |b|$$

if $R = |b| < 1$, then

$$\int_0^1 f(\hat{w}_i) \mathrm{d}\hat{w}_i = 1$$

Appendix A.4. Proof of Theorem 4

$$E(\hat{w}_i^r) = \int_0^1 \hat{w}_i^r f(\hat{w}_i) \mathrm{d}\hat{w}_i$$
$$= (1-b)^{m_2/2} \sum_{j=0}^{\infty} \left[C_j^{(m+j)-1} b^j \frac{B((m_2/2)+r+j,\, m_1/2)}{B(m_2/2,\, m_1/2)} \right]$$

Again, by the previous theorem (Apostol, 1974, p. 193), with

$$R_r = \limsup_{n\to\infty} \left| \frac{(m+n)(m_2+2r+2n)}{(1+n)(2m+2r+2n)} b \right| = |b|$$

$R_r = |b| < 1$, then the rth-order raw moment of $\hat{w}_i$ exists.

Appendix A.5. Proof of Corollary 1

Let

$$B_j^r = \frac{B((m_2/2)+r+j,\, m_2/2)}{B(m_2/2,\, m_1/2)}$$
$$B_j^{r+1} = \frac{m_2+2r+2j}{m_1+m_2+2r+2j} B_j^r$$

Then

$$B_j^{r+1} < B_j^r, \forall r \in \mathbf{N}$$

and therefore

$$E(\hat{w}_i^{r+1}) = (1-b)^{m_2/2} \sum_{j=0}^{\infty} C_j^{(m+j)-1} b^j B_j^{r+1}$$
$$< (1-b)^{m_2/2} \sum_{j=0}^{\infty} C_j^{(m+j)-1} b^j B_j^{r} = E(\hat{w}^r), \forall r \in \mathbf{N}$$

Appendix A.6. Pearson Type I Approximation

We must compute some important coefficients such as

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \beta_2 = \frac{\mu_4}{\mu_2^2}$$

then check the following conditions

$$6 + 3\beta_1 - 2\beta_2 > 0, \ \beta_2 - \beta_1 - 1 > 0$$

and the interpolation between x_0 and x_1 is

$$x(\theta) = (1-\theta)x_0 + \theta x_1 - \frac{1}{4}\theta(1-\theta)[\Delta^2 x_0 + \Delta^2 x_1]$$

References

Apostol, T.M. (1974) *Mathematical Analysis* (New York: Wiley).

Bunn, D.W. (1986) Statistical efficiency in the linear combination of forecasts, *International Journal of Forecasting*, 1, pp. 151–163.

Dickinson, J.P. (1973) Some statistical results in the combination of forecasts, *Operational Research Quarterly*, 24, pp. 253–260.

Johnson, N.L., Nixon, E., Amos, D.E. & Pearson, E.S. (1963) Table of percentage points of Pearson curves, for given $\sqrt{\beta_1}$ and β_2 expressed in standard measure, *Biometrika*, 50, pp. 459–498.

Lee, J.C. & Hu, L. (1996) On the distribution of linear functions of independent F and U variates, *Statistics & Probability Letters*, 26, pp. 339–346.

Liang, K.Y. (1992) On the sign of the optimal combining weights under the error-variance minimizing criterion, *Journal of Forecasting*, 11, pp. 719–723.

Newbold, P. & Granger, C.W.J. (1974) Experience with forecasting univariate time series and combination of forecasts, *Journal of Royal Statistics Society, Series A*, pp. 131–146.

Reid, D.J. (1969) A comparative study of time-series prediction techniques on economic data. PhD thesis, Department of Mathematics, University of Nottingham.

Winkler, R.L. & Clemen, R.T. (1992) Sensitivity of weights in combining forecasts, *Operations Research*, 40, pp. 609–614.