Theoretical Economics Letters, 2012, 2, 435-437
http://dx.doi.org/10.4236/tel.2012.25080 Published Online December 2012 (http://www.SciRP.org/journal/tel)

Some Exact Results for an Asset Pricing Test Based on the Average *F* Distribution

Soosung Hwang[1], Stephen E. Satchell[2,3]
[1]School of Economics, Sungkyunkwan University, Seoul, South Korea
[2]Trinity College, University of Cambridge, Cambridge, UK
[3]Finance Department, Sydney University, Sydney, Australia
Email: shwang@skku.edu, ses11@econ.cam.ac.uk

Received September 11, 2012; revised October 12, 2012; accepted November 15, 2012

ABSTRACT

We provide some exact results for an asset pricing theory test statistic based on the average *F* distribution. This test is preferred to existing procedures because it deals with the case of more assets than data points. The case mentioned is the practical one that asset managers routinely have to consider.

Keywords: Average *F* distribution; Asset Pricing

1. Introduction

The idea of the average *F* test was first introduced to the literature by [1] as a means of testing asset pricing theories in linear factor models. Recently [2] developed the idea further by focusing on the *average* pricing error, extending the multivariate *F* test of [3]. They show that the average *F* test can be applied to thousands of individual stocks rather than a smaller number of portfolios and thus does not suffer from the information loss or the data snooping biases. In addition, the test is robust to ellipticity. More importantly, [2] demonstrate that the power of average *F* test continues to increase as the number of stocks increases.

One drawback of the average *F* test is that [2] did not provide the closed form solution for the average *F* density function. Despite the fact that the average *F* statistic has been used in other areas of econometrics, e.g., [4] in the study of structural breaks of unknown timing in regression models, the functional form of the average *F* distribution remains unknown.

In this study we propose a few analytical developments for the average *F* distribution. Although the complete functional form is not provided, our results might be useful toward further research in the future.

2. Definition of the Average *F* Distribution

A testable version of linear factor models is

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\boldsymbol{r}_t^f + \boldsymbol{\varepsilon}_t, \tag{1}$$

where $\boldsymbol{r}_t$ is a $(N \times 1)$ vector of excess returns for N assets and $\boldsymbol{r}_t^f$ is a $(K \times 1)$ vector of factor portfolio returns, $\boldsymbol{\alpha} \equiv (\alpha_1, \alpha_2, \cdots, \alpha_N)'$ is a vector of intercepts, $\boldsymbol{\beta} \equiv (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_N)$ is an $(N \times K)$ matrix of factor sensitivities, and $\varepsilon_t \equiv (\varepsilon_{1,t}\varepsilon_{2,t}, \cdots, \varepsilon_{N,t})'$ is a $(N \times 1)$ vector of idiosyncratic errorswhose covariance matrix is $E(\varepsilon_t \varepsilon_t') = \boldsymbol{\Sigma}$. For the null hypothesis $H_0^a : \boldsymbol{\alpha} = 0$ tested against the alternative hypothesis $H_1^a : \boldsymbol{\alpha} \neq 0$, the average F-test statistic is defined as

$$S = \frac{Tc}{N} \sum_{n=1}^{N} \frac{\hat{\alpha}_n^2}{\hat{\sigma}_n^2} \tag{2}$$

where

$$\hat{\sigma}_n^2 = \sum_{t=1}^{T} \left(r_{n,t} - \hat{\alpha}_n - \hat{\boldsymbol{\beta}}_n \boldsymbol{r}_t^f \right)^2 \Big/ (T - K - 1)$$

$$c = \left(1 + \hat{\boldsymbol{\mu}}_K' \hat{\boldsymbol{\Omega}}_K^{-1} \hat{\boldsymbol{\mu}}_K \right)^{-1},$$

$$\hat{\boldsymbol{\Omega}}_K = \frac{1}{\mathrm{T}} \sum_{t=1}^{T} \left(\boldsymbol{r}_{\mathrm{t}}^{\mathrm{f}} - \hat{\boldsymbol{\mu}}_K \right) \left(\boldsymbol{r}_{\mathrm{t}}^{\mathrm{f}} - \hat{\boldsymbol{\mu}}_K \right)',$$

and $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ are the maximum likelihood estimators of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. Under the classical assumption that asset returns are multivariate normal conditional on factors, the average *F* statistic is distributed as

$$S \sim \frac{1}{N} \sum_{n=1}^{N} F_n(1, T - K - 1), \tag{3}$$

where $F_n(1, T - K - 1)$ is a F statistic with 1 degree of freedom in the numerator and $(T - K - 1)$ degrees of freedom in the denominator.

3. Characteristic Function of the Average *F* Distribution

The distribution function of the average F statistic is unknown. Note that all F-distributions in Equation (3) have the same degrees of freedom, and S is thus distributed as the sample mean of N independent and identically distributed F distributions. Let x be a variable distributed as $F(1,n)$, where $n \equiv T-K-1$, and denote its probability density function as $pdf(x)$. Then the characteristic function of the $F(1,n)$ distribution can be derived as follows

$$\Phi(t,n) = \int_0^\infty pdf(x)\mathrm{e}^{itx}\mathrm{d}x = \int_0^\infty \frac{\Gamma\left(\frac{1}{2}(1+n)\right)n^{n/2}}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}n\right)} \frac{x^{-1/2}}{(n+x)^{\frac{1+n}{2}}}\mathrm{e}^{itx}\mathrm{d}x \tag{4}$$

Let $y = x/n$ and $n\mathrm{d}y = \mathrm{d}x$, then

$$\begin{aligned}\Phi(t,n) &= \frac{\Gamma\left(\frac{1}{2}(1+n)\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}n\right)}\int_0^\infty (1+y)^{-\frac{1+n}{2}} y^{-\frac{1}{2}}\mathrm{e}^{itny}\mathrm{d}y \\ &= \frac{\Gamma\left(\frac{1}{2}(1+n)\right)}{\Gamma\left(\frac{1}{2}n\right)}\Psi\left(\frac{1}{2},1-\frac{1}{2}n;-nit\right),\end{aligned} \tag{5}$$

where $\Gamma(\cdot)$ is the gamma function, i is the imaginary number, and $\Psi(\cdot)$ is Tricomi's confluent hypergeometric function. Equation (5) was first formulated by [5]. Tricomi's confluent hypergeometric function is

$$\begin{aligned}&\Psi\left(\frac{1}{2},1-\frac{1}{2}n;-nit\right) \\ &= \frac{\Gamma\left(\frac{1}{2}n\right)}{\Gamma\left(\frac{1}{2}(1+n)\right)}\,{}_1F_1\left(\frac{1}{2},1-\frac{1}{2}n;-nit\right) \\ &+ \frac{\Gamma\left(-\frac{1}{2}n\right)}{\Gamma\left(\frac{1}{2}\right)}(-nit)^{n/2}\,{}_1F_1\left(\frac{1}{2}+\frac{1}{2}n,1+\frac{1}{2}n;-nit\right),\end{aligned} \tag{6}$$

where ${}_1F_1(\cdot)$ is Kummer's confluent hypergeometric function which is defined as

$$\begin{aligned}{}_1F_1(a,b;z) &= 1+\frac{a}{b}\frac{z}{1!}+\frac{a(a+1)}{b(b+1)}\frac{z^2}{2!} \\ &+\frac{a(a+1)(a+2)}{b(b+1)(b+2)}\frac{z^3}{3!}+\cdots\end{aligned} \tag{7}$$

See [6] for a detailed explanation of various types of hypergeometric functions and their applications to economic theory.

If b in Equation (7) is a non-positive integer, ${}_1F_1(a,b;z)$ and thus

$$\Psi\left(\frac{1}{2},1-\frac{1}{2}n;-nit\right)$$

is not defined. Note that n is a positive integer as it represents the degrees of freedom in the denominator of the $F(1,n)$ distribution; thus, we need

$$1-\frac{1}{2}n \text{ and } 1+\frac{1}{2}n$$

in Equation (6) to be positive integers. However, since n is a positive integer, both

$$1-\frac{1}{2}n \text{ and } 1+\frac{1}{2}n$$

cannot be kept to be positive integers. More generally,

$$\text{when } \frac{n}{2}\in Z,$$

we have a definition referred to as the "logarithmic case" alternative to Tricomi's confluent hypergeometric function in (6). See [6] and [7] (Vol. 1, pp. 260-262 and Vol. 2, p. 9) for discussions on the logarithmic case.

Let $\phi_j(t,n)$ be defined as the characteristic function of the j^{th} independent $F(1,n)$ variable. Then, the characteristic function of S is

$$\Phi(t,n) = \prod_{j=1}^N \phi_j\left(\frac{t}{N},n\right) = \left[\phi\left(\frac{t}{N},n\right)\right]^N, \tag{8}$$

where $\phi(\cdot)$ is defined in (5). Therefore, the density function of the average F statistic S, $pdf(y)$, under the null hypothesis is obtained by the following;

$$pdf(y) = \frac{1}{(2\pi)}\int\left[\phi\left(\frac{t}{N},n\right)\right]^N \mathrm{e}^{-ity}\mathrm{d}t, \tag{9}$$

where y is a variable distributed as the average of the N different $F(1,n)$ distributions. This mean of F-distributions can be used when the variance-covariance matrix $\mathbf{\Sigma}$ is a diagonal matrix.

4. The Exact Distribution of Average *F* Test for Small *N*

When $N=1$, we have $S \sim F(1,T-K-1)$. Using the result that $F(1,T-K-1)$ is the square of $t(T-K-1)$, *i.e.*, a t distribution with $T-K-1$ degrees of freedom, we see that the pdf of S is given by $pdf_1(y)$, letting $v = T-K-1 > 0$,

$$pdf_1(y) = y^{-1/2}\left(1+\frac{y}{v}\right)^{-\left(\frac{v+1}{2}\right)}\frac{1}{\sqrt{v}B\left(\frac{1}{2},\frac{v}{2}\right)}, \tag{10}$$

where the pdf of $t(T-K-1)$ can be found in [8].

To find the pdf of S when $N=2$, $pdf_2(y)$, we proceed as follows. Let Y_2 be the associated random variable, and Y_{11} and Y_{12} be the two independent

$F(1, T-K-1)$ variables. Then we have

$$\Pr(Y_2 \le y) = \Pr(Y_{11} + Y_{12} < 2y) = \int_0^{2y} \Pr(Y_{11} \le 2y - x) pdf_1(x) \mathrm{d}x.$$

Therefore $pdf_2(x) = 2\int_0^{2y} pdf_1(2y - x) pdf_1(x) \mathrm{d}x$, where $pdf_1(x)$ is given by Equation (10). More generally, by induction, it follows that

$$pdf_k(y) = k \int_0^{ky} pdf_{k-1}(ky - x) pdf_1(x) \mathrm{d}x. \quad (11)$$

Although it is hard to make much progress with Equation (11) in obtaining closed form solutions, we note the following. From known moments of the t distribution, it is possible to calculate the moments of S for any N, where they exist.

Proposition 1. *The moments of S exist for*

$$s < \frac{T-K-1}{2}.$$

Proof. Let $S = \frac{1}{N}\sum_{i=1}^{N} y_i$. Then $S^s = \frac{1}{N^s}\left(\sum_{i=1}^{N} y_i\right)^s$ so that the highest order term, for any i, is y_i^s. Now from Equation (10),

$$c = \frac{1}{\sqrt{v} B\left(\frac{1}{2}, \frac{v}{2}\right)},$$

and thus $E(y_i^s) = c\int_0^{\infty} y_i^{s-\frac{1}{2}} \left(1 + \frac{y_i}{v}\right)^{-\left(\frac{v+1}{2}\right)} \mathrm{d}y_i$ which exists if $s - \frac{1}{2} - \frac{v+1}{2} < -1$, or $s < \frac{v}{2}$.

Proposition 2 *For* $N = 1$, $pdf_1(y)$ *can be represented as a scale Beta type II function.*

Proof. For $pdf_1(y)$ given by Equation (10), let

$$z = \frac{1}{1 + \frac{y}{v}}.$$

Then $\mathrm{d}y = \frac{v}{z^2}\mathrm{d}z$ and simple change of variable shows that z is a Beta $\left(\frac{v}{2}, \frac{1}{2}\right)$ random variable.

Since Proposition 2 establishes that $pdf_1(y)$ is a scaled Beta, we now have a representation of S. Denoting $S = S(N)$ to reflect the dependence on N, it follows from Proposition 2 that

$$\mathrm{S}(N) = \frac{1}{N}\sum_{j=1}^{N} \frac{1}{v} B_2\left(\frac{v}{2}, \frac{1}{2}\right) \quad (12)$$

where $B_2\left(\frac{v}{2}, \frac{1}{2}\right)$ denotes a type II beta with parameters $\frac{v}{2}$ and $\frac{1}{2}$ and the v outside $B_2\left(\frac{v}{2}, \frac{1}{2}\right)$ reflects the scale factors. Thus Equation (12) establishes that $S(N)$ can be represented as a linear combination of Beta type II distributions.

The literature on density functions of linear combination of Beta distributions is rather sparse. [9] present expressions for linear combinations of Beta distributions when $N = 2$. Thus using their results we can arrive at an expression for $pdf_2(y)$ which is complex and depends upon hypergeometric functions. Extensions for $N > 2$ do not appear to be derived as yet.

5. Conclusion

We provide some developments on the average F test distribution. Although simulation of the statistic is straightforward, an understanding of the functional form is invaluable in terms of appreciation of the properties of the test statistic. We leave a full solution of the problem for future study.

REFERENCES

[1] J. Affleck-Graves and B. McDonald, "Multivariate Tests of Asset Pricing: The Comparative Power of Alternative Statistics," *Journal of Financial and Quantitative Analysis*, Vol. 25, No. 2, 1990, pp. 163-185. doi:10.2307/2330822

[2] S. Hwang and S. E. Satchell, "Testing Linear Factor Models on Individual Stocks Using the Average F Test," *European Journal of Finance*, Forthcoming. doi:10.1080/1351847X.2012.717097

[3] M. Gibbons, S. Ross and J. Shanken, "A Test of the Efficiency of a Given Portfolio," *Econometrica*, Vol. 57, No. 5, 1989, pp. 1121-1152. doi:10.2307/1913625

[4] D. W. K. Andrew and W. Ploberger, "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative," *Econometrica*, Vol. 62, No. 6, 1994, pp. 1383-1414. doi:10.2307/2951753

[5] P. C. B. Phillips, "The True Characteristic Function of the F Distribution," *Biometrika*, Vol. 69, No. 1, 1982, pp. 261-264.

[6] K. M. Abadir, "An Introduction to Hypergeometirc Functions for Economists," *Econometric Review*, Vol. 18, No. 3, 1999, pp. 287-330. doi:10.1080/07474939908800447

[7] A. Erdélyi, "Higher Transcendental Functions," Mc. Graw-Hill, New York, 1953.

[8] A. Stuart and K. Ord, "Kendall's Advanced Theory of Statistics, Volume 1 Distribution Theory," 6th Edition, Edward Arnold, London, 1994.

[9] T. Pham-Gia and N. Turkkan, "Distribution of the Linear Combination of Two General Beta Variables and Applications," *Communication in Statistics*; *Theory and Methods*, Vol. 27, No. 7, 1998, pp. 1851-1869. doi:10.1080/03610929808832194

