THE JOURNAL OF FINANCE · VOL. XXXIII, NO. 5 · DECEMBER 1978

"ARE BETAS BEST?"†

EDWIN J. ELTON,* MARTIN J. GRUBER* AND THOMAS J. URICH**

IN 1973 ELTON AND GRUBER (3) published an article in which they investigated the accuracy of forecasts of the correlation structure between securities produced by a group of alternative models. As pointed out in that article, while analysts may be capable of providing estimates of returns and variances, the development of estimates of correlation coefficients from anything other than models utilizing historical data is highly unlikely. The article after investigating the accuracy of certain forecasting techniques suggested that those techniques which provided the best forecasts could lead to a simplification of the portfolio problem. Elton, Gruber and Padberg in a series of subsequent articles (4), (5) and (6) have shown that if one accepts any of the assumptions underlying the models which worked best as forecasters in (3) then the optimal portfolio can be determined quickly and easily without the use of a computer.

In this paper we shall re-examine the forecasting ability of some of the techniques which provided the best set of forecasts in (3) as well as some new variations of these techniques. This re-examination is motivated by two factors. The first is that development of simple rules for optimal portfolio selection (4), (5), and (6) means that the likelihood of one or more of these techniques being used has been substantially increased, thus the choice between them becomes much more critical. The second is that some improvements in estimating the basic techniques have been reported in the literature and have gained acceptance in industry and it is worthwhile to compare the improved version of the basic models.

I. TECHNIQUES

This section describes the alternative types of models we used for estimating future correlation coefficients: a historical model, single index models and a constant correlation model.

A. *A Full Historical Model*

The simplest method of estimating future correlation coefficients using historical data is to assume that past values of these coefficients are the best estimates of their future values. In short calculate each pairwise correlation coefficient over a historical period and use this value as an estimate of the future. No assumption is made as to how or why securities move together. Instead, the amount of their co-movement is estimated directly. Several studies [e.g., (2)&(3)] have shown that while there is useful information contained in historical

* Graduate School of Business Administration New York University

** Brauch College, City University of New York

† The authors would like to thank the Institute for Quantitative Research in Finance for Financial Support

correlation coefficients, other techniques perform better. We have used the historical matrix in this study because it is common to all studies and can usefully serve as a benchmark against which other techniques can be evaluated. Historical correlation coefficients were calculated using the five years of data prior to the forecast period.

B. *Single Index Models*

The single index model is the most widely used method of estimating correlation coefficients. As is well known the key assumption underlying its use is that securities systematically move together only because of a common response to a single index. We've used four variants of this model: unadjusted Betas, Vasicek adjustment to Betas, Blume's Beta adjustment and an assumed Beta of one for all stocks.

Unadjusted Betas are the Betas obtained from a least square regression of security returns on a market index during some historical period. These Betas are then used to estimate future correlation coefficients. Throughout this study we use as an index the S & P composite index adjusted for dividends. Unadjusted Betas were obtained by fitting a regression using data from the 5 years prior to the forecast period.

The Vasicek (See (11) for a detailed discussion) procedure consists of taking a weighted average of the unadjusted Beta and the average Beta for the sample of stocks under consideration. The weights add to one and depend on the standard error of the individual Beta and the standard deviation of the distribution of Betas in the sample. Once again the parameters were obtained by using data from the 5 years prior to the forecast period.

The Blume adjustment procedure consists of regressing Betas from one historical period on Betas from a prior period and then using this regression to adjust Betas for the forecast period (see Blume (1) for a detailed discussion of this procedure). Consider three sequential five year periods of data. Unadjusted Betas from the second period are regressed on unadjusted Betas from the first period. Unadjusted Betas from the second period are then substituted into this regression, as the independent variable to obtain an adjusted estimate of the Beta in the third period.

The final estimate of Beta was obtained by assuming that all Betas are one. The use of a Beta of one was motivated by Fisher and Kamin's (8) finding that this naive forecast performed surprisingly well compared to some more sophisticated techniques. Thus we have four estimates of future correlation coefficients using single index models.

Using standard symbols, estimates of correlation coefficients are obtained from Betas by noting that

$$\mathrm{Cov}(ij) = \rho_{ij}\ \sigma_i\sigma_j = \beta_i\beta_j\sigma_m{}^2$$

thus

$$\rho_{ij} = \frac{\beta_i\beta_j\sigma_m{}^2}{\sigma_i\sigma_j}$$

C. *Constant Correlation Model*

An alternative to the single index models just presented is to assume that historical data only contains information concerning the mean correlation coefficients and that observed pair-wise differences from the average are random or sufficiently unstable, so that zero is a better estimate (than their historical level) of their future value. The most aggregate averaging possible is to set every correlation coefficient equal to the average of all correlation coefficients. In an earlier study we examined this technique primarily as a naive model against which to compare more sophisticated models. To our surprise it was among the models that produced the best forecasts, outperforming the unadjusted Beta and full historical model. Hence it is well worth continuing to examine it. We call this technique Overall Mean.

II. Empirical Tests

The superiority of each of the techniques under consideration is judged by its ability to forecast the correlation matrix of security returns in a period subsequent to that over which the technique is fitted. Forecast accuracy will be judged over 5 year periods. In order to insure the robustness of our techniques, forecasts will be made for each of two 100 firm samples (as well as the combined sample) over two 5 year non-overlapping periods.[1] The first set of forecasts are prepared using data from 1956 to 1965 for the Blume technique and 1961 to 1965 for all other techniques and the forecasts are judged for accuracy against the actual correlation matrix for the years 1966 to 1970. For the second set of tests the parallel dates are 1961–1970, 1966–1970, and 1971–1975. An additional 5 year period is required for the Blume technique to parameterize the model.

Two types of tests were employed to judge differences in the forecasts produced by alternative techniques. First we examined the differences in the absolute forecast error between each pair of techniques. We judged one technique dominated the second, if the mean of these differences was, statistically, significantly different from zero. To be more precise, for each correlation coefficient, we computed the absolute value of the difference between the expost correlation coefficient and the forecast correlation coefficient produced by each technique. Then, taking the techniques two at a time (say technique A and B), the difference in the absolute values of the forecast errors was computed for each correlation coefficient. This allowed us to compute the distribution of the difference in the absolute value of the forecast error between technique A and B across all firms as well as the mean and standard deviation of this distribution. Technique A was said to dominate technique B if the mean of this distribution was negative and significantly different from zero at the 5% level.[2]

The second test we employed was to examine the distribution of the differ-

1. To be more precise, we made forecasts on the basis of the 200 firm sample and we observed how well these forecasts did over two 100 firm subsamples as well as over the 200 firm sample.

2. From the central limit theorem, the distribution of the mean of these frequency functions is normally distributed with mean equal to the mean of the frequency function. Thus standard tests for means different from zero can be employed.

ences in absolute forecast errors and see if the probability of any size error was less with one technique then a second. We examined 10 intervals of .02. For example consider two of our techniques Overall Mean and Full Historical. We first examined the probability that overall mean would have an error greater than .20 larger than full historical. We then examined the probability that Full Historical would have an error .20 larger than overall mean. We continued this for .18 larger, .16 larger etc. If the odds of all size errors were consistently less for one technique than a second, then that technique dominates independent of the individuals loss function.[3] This test is a much more stringent test of dominance than the one discussed above. In some cases the mean error was smaller for one technique compared to a second and the difference was statistically significant but the first technique tended to have many large errors.

In examining the forecast accuracy of alternative models we felt it was worthwhile to look at the results both when the average forecast from each technique was adjusted to have the same value and when it was not. The forecast errors from each model are due in part to the error in forecasting the mean correlation coefficient and in part due to errors in forecasting differences from the mean. Several of the techniques can be expected to produce biased estimates of the means (as will be discussed in detail in section IV). Since the individual employing these techniques can easily adjust for differences in mean forecasts it is worthwhile examining forecast accuracy both with and without the mean adjustment. In the next section we examine unadjusted forecasts.

III. Unadjusted Forecasts

As shown in Table 1, in each five year period, Overall Mean had a lower error at a statistically significant level than any other technique. In addition, as shown in Tables 2 and 3, for both time periods the probability of obtaining an error of any size is always smaller with the Overall Mean than with any other procedure except Blume. Overall mean dominates Blume for 9 of the 10 categories of error size in each time period. As another check on the dominance of overall mean we split each 200 firm sample into two 100 firm samples. Once again with one exception overall mean dominates all other techniques at a statistically significant level. The one exception is that in one sample for one time period, while

TABLE 1

Average Absolute Error*
(No Adjustment)

First Five Years		Second Five Years		Combined	
1. Overall Mean	.1169	1. Overall Mean	.1415	1. Overall Mean	.1292
2. Blume Beta	.1270	2. Blume Beta	.1499	2. Blume Beta	.1385
3. Vasicek Beta	.1289	⌈3. Unadjusted Beta	.1539	3. Vasicek Beta	.1419
4. Unadjusted Beta	.1348	| 4. Full Historical	.1545	4. Unadjusted Beta	.1444
5. Beta = 1	.1378	⌊5. Vasicek	.1548	5. Full Historical	.1491
6. Full Historical	.1436	6. Beta = 1	.1776	6. Beta = 1	.1577

* All differences are statistically significant unless grouped by a bracket.

3. This follows directly from the principles of stochastic dominance.

TABLE 2

DOMINANCE BY CUMULATIVE FREQUENCY FUNCTION
(Unadjusted)
First Five Years

	Full Historical	Beta = 1	Unadjusted Beta	Vasicek Beta	Blume Beta	Overall Mean
Full Historical	X					
Beta = 1	10	X				
Unadjusted Beta	10	8	X			
Vasicek Beta	10	8	5	X		
Blume Beta	10	10	10	6	X	
Overall Mean	10	10	10	10	9	X

The number in the table indicates the number of times the technique on the left had a larger cumulative frequency function (and hence small errors) than the technique on the top. For example, the 10 in the cell next to Overall Mean indicates the odds of having an error less than any amount are always higher with Overall Mean than with Full Historical.

Overall Mean produced a smaller error, the difference is not statistically significant.

The second best performing technique is the Blume procedure. It dominated all other techniques in terms of mean absolute error and all but the Vasicek procedure in terms of distributional properties. The dominance of the Blume procedure carried over to the subsamples. The Blume technique performed better than all other techniques and at a significant level except for the Vasicek technique in one subsample and the full historical in a second.

The next best performing technique was the Vasicek adjustment to Beta. However, its ranking varied in the two five year periods. This variation was carried over in the subsamples.

The ranking among the remaining techniques also varied somewhat. In general the remaining techniques were (from best to worst) the unadjusted Beta, full historical, and Beta = 1.

TABLE 3

DOMINANCE BY CUMULATIVE FREQUENCY FUNCTION
(Unadjusted)
Second Five Years

	Beta = 1	Vasicek Beta	Full Historical	Unadjusted Beta	Blume Beta	Overall Mean
Beta = 1	X					
Vasicek Beta	10	X				
Full Historical	10	3	X			
Unadjusted Beta	10	9	7	X		
Blume Beta	10	10	10	10	X	
Overall Mean	10	10	10	10	9	X

* See note to Table 2.

IV. Adjustment by the Mean Correlation Coefficient

Almost all of the models presented in this paper lead to different estimates of the average future correlation coefficient. The one exception is that the Overall Mean and the Full Historical technique give the same estimate. Thus any technique might appear superior (inferior) because of differences in estimating the mean correlation coefficient or because of differences in the estimation of the variation of each correlation coefficient from the mean. Alternative techniques will produce estimates of the future mean correlation coefficient that are either lower than or consistently higher than the historical mean correlation coefficient. Unless there is a reason to suspect a drift in the average correlation coefficient overtime, it is useful to judge the performance of any technique with the effect of bias in its mean forecast removed. We shall do so shortly but first we will discuss why each technique is likely to produce estimates of future correlation coefficients that tend to be above or below the average correlation in the sample on which the technique is fitted.

Table 4 presents the average correlation coefficient estimated by each of the forecasting techniques for each of the samples under study. Note that the mean estimate for the Full Historical Technique (and the Overall Mean Technique) is equal to the true mean from the sample.

The average correlation coefficient estimated by the Unadjusted Beta Method is smaller than the average correlation that existed over the period of time to which the model was fitted. The same result was found in Elton and Gruber (3) though a different sample of firms was employed over two different time periods. All of the Beta methods assume that the only correlation between securities arises because of common correlation with the market. To the extent that there are additional sources of correlation (e.g. industry affects) present, ignoring them will lead to downward biased estimates of the correlation coefficient.[4]

Yet despite the fact that the Beta technique ignores all sources of correlation except that associated with the market, the Blume technique produces higher estimates of the correlation coefficient than the sample to which it was fitted. There are two factors that account for this. One is that the average Beta in the sample was increasing in the periods under study. The Blume technique results in an extrapolation of this rise in Beta and thus for the period studied in higher

TABLE 4

Forecast of Average Correlation Coefficient

First Five Years		Second Five Years	
Blume Beta	.354	Blume Beta	.356
Full Historical	.302	Full Historical	.334
Overall Mean	.302	Overall Mean	.334
Beta = 1	.290	Beta = 1	.305
Unadjusted Beta	.276	Unadjusted Beta	.302
Vasicek Beta	.275	Vasicek Beta	.298
Realization	.334	Realization	.359

4. Additional sources of correlation have been discussed and empirically demonstrated in (7) and (9).

estimates of Betas (and correlations) than the unadjusted Betas. This is easily seen from the fact that the Blume technique produced average Betas of 1.14 and 1.14 for the two periods while the unadjusted Betas for these two periods was 1.04 and 1.11. The second reason the Blume Betas result in higher correlations is because of the adjustment of all Betas towards the mean. Remember that the correlation coefficients depend on the product of the Betas for alternative securities. The average product of two Betas with the same mean but less dispersion around the mean will be higher than the average product of the unadjusted distribution.

The second (not the first) of these influences will also be present in the Vasicek technique. Why then does the Vasicek adjustment procedure result in lower Betas? The bias comes from the weighting procedure used to obtain the adjusted Beta. The adjusted Beta is a combination of the stocks own Beta and the average Beta in the sample. The weight placed on the stocks own Beta depends on the inverse of the variance of the stocks Beta. Since high Beta stocks tend to have high standard errors in their Beta estimate while low Beta stocks have low standard errors this procedure results in a large decrease in the Betas for high Beta stocks relative to the increase in the Beta for low Beta stocks. The net result is an estimate of average Beta below its unadjusted level and a low estimate of average correlation coefficient. For example, in the two periods under study, the average Beta from the Vasicek technique was 1.02 and 1.07 which is much lower than the estimates produced by the unadjusted Beta method or the Blume method.

The remaining method is to set all Betas equal to one.[5] This technique led to low estimates of the mean correlation coefficient because the average Betas in our two samples were 1.04 and 1.11.

Unless there is some reason why one of these techniques captures a non random movement in the mean correlation coefficient, and the authors see no reason to suspect this is true, it is worthwhile to examine the forecast ability of these models when the forecasts are forced to have the same mean. We can consider this the ability of each model to forecast correlation coefficients when the forecast of the mean correlation coefficient is done exogenously. In particular two forecasts of the average future correlation coefficient will be tried. The first will assume that the average correlation coefficient follows a zero drift random walk. If this is true then the optimal forecast is the last observed average correlation coefficient.[6,7] Our second forecasting method will assume perfect knowledge and will adjust by the actual average correlation coefficient in the future period. In both cases each correlation coefficient estimated by some technique is adjusted by subtracting from it the mean correlation coefficient for that technique and adding the new estimate of the mean.

The results of adjusting all forecasts for the historic average correlation

5. The estimates from this technique might have improved if the Betas were set equal to the average in the sample rather then one. However given the poor performance of this technique this change is unlikely to effect any conclusions.

6. Since this number can easily be computed from historical data if this adjustment leads to better forecasts the user can easily incorporate the adjustment in his forecasting scheme.

7. By optimal forecast we mean unbiased since an individual's loss function could result in an optimal forecast that is biased.

coefficient are presented in Table 5. The average absolute error of the Overall Mean and Full Historical technique are unchanged since their average forecast is unchanged. The average absolute error of every other technique except Blume Betas goes down in both periods. Blume does better on an unadjusted bases because it is upward biased and there was a continued upward movement in correlation coefficients over the period examined. The most striking result from Table 5 is that the Overall Mean continues to dominate all other techniques at a statistically significant level in both periods. This is true not only using the mean absolute error but also using the same cumulative frequency function as employed in Part I. Using this technique the Overall Mean dominates each other technique 10 out of 10 possible times. Thus, irrespective of an individuals loss function, the Overall Mean dominates.

Another interesting point is the improvement in the Vasicek technique. Correcting for its downward bias in estimating the mean made it the second best forecasting technique. It dominates all other techniques (but Overall Mean) at a statistically significant level in both periods. As mentioned above the Blume technique now performs worse than it did without the mean adjustment and is virtually indistinguishable from the unadjusted Beta technique. The fact that setting Beta equal to one performs worse than any other Beta technique when the mean adjustment is made indicates that differences between Betas, whether adjusted or not, convey real information about the future correlation structure. Finally the poor performance of the Full Historical correlation matrix indicates that pairwise correlation coefficients contain a lot of random noise. The Overall Mean and Beta techniques can be viewed as smoothing techniques, with the Overall Mean method involving the most stringent smoothing. The ordering of the techniques indicates that smoothing across correlation coefficients eliminates random noise and improves forecasting ability.

It is worth while examining how robust the relative performance of these techniques is across non-overlaping subsamples of firms. As mentioned earlier we prepared forecasts for each technique for each subsample in each of two periods. The rankings for each technique are shown in Table 6. Notice that the rankings are quite consistent between subsamples as well as for the overall sample. Overall Mean performs best in each subsample. The Vasicek Beta technique performs second best in three of the four subsamples. The one time it is third rather than second its difference from the second technique is not

TABLE 5

AVERAGE ABSOLUTE ERRORS*
(Historical Adjustment)

First Five Years		Second Five Years		Combined	
1. Overall Mean	.1169	1. Overall Mean	.1415	1. Overall Mean	.1292
2. Vasicek Beta	.1231	2. Vasicek Beta	.1458	2. Vasicek Beta	.1345
3. Blume Beta	.1275	3. Unadjusted Beta	.1468	3. Unadjusted Beta	.1372
4. Unadjusted Beta	.1293	4. Blume Beta	.1522	4. Blume Beta	.1399
5. Beta = 1	.1358	5. Full Historical	.1545	5. Full Historical	.1491
6. Full Historical	.1436	6. Beta = 1	.1720	6. Beta = 1	.1539

* All differences are statistically significant unless grouped by a bracket.

TABLE 6

RANK OF EACH TECHNIQUE IN EACH SUBSAMPLE
(Rank on Mean Absolute Error)
historical adjustment

Technique	First Five Years Sample 1	First Five Years Sample 2	Second Five Years Sample 1	Second Five Years Sample 2
Overall Mean	1	1	1	1
Vasicek Beta	2	2	3*	2
Unadjusted Beta	3	4	2*	4*
Blume Beta	4	3	4	3*
Full Historical	6*	6*	5	5
Beta = 1	5*	5*	6	6

* Indicates that the two entries in the column were not different at a statistically significant level.

statistically significant. The next two techniques in performance are the Blume Beta and the Unadjusted Beta. They switch back and forth in ranking but are worse than the two previously mentioned techniques and better than the two remaining techniques. The two worst techniques are the Full Historical technique and the Beta equals one model. They are virtually indistinguishable from each other.

Table 7 presents the results of forecasting future correlation coefficients when the techniques are adjusted so that the average correlation coefficient predicted by each model is equal to the true future average correlation coefficient. Obviously this leads to an improvement in forecast accuracy. Of more interest is the fact that the ordering and statistical significance of results is identical to that obtained when the adjustment is made using the historic average correlation coefficient.

V. CONCLUSION

In this paper we demonstrated that Overall Mean is a prefered method of forecasting future correlation coefficients in comparison to the best of the Beta time series techniques. In addition we discussed the bias, in the mean estimate produced by standard Beta estimating techniques. Correcting for this bias had no

TABLE 7

AVERAGE ABSOLUTE ERROR*
(Future Adjustment)

First Five Years		Second Five Years		Combined	
1. Overall Mean	.1145	1. Overall Mean	.1384	1. Overall Mean	.1265
2. Vasicek Beta	.1211	2. Vasicek Beta	.1432	2. Vasicek Beta	.1322
3. Blume Beta	.1256	3. Unadjusted Beta	.1443	3. Unadjusted Beta	.1354
4. Unadjusted Beta	.1275	4. Blume Beta	.1497	4. Blume Beta	.1377
5. Beta = 1	.1335	5. Full Historical	.1521	5. Full Historical	.1416
6. Full Historical	.1414	6. Beta = 1	.1696	6. Beta = 1	.1516

* All differences are statistically significant unless grouped by a bracket.

effect on the dominance of Overall Mean but did change the ranking within the Beta techniques.

REFERENCES

1 Marshal Blume. "On The Assessment of Risk", *Journal of Finance* (March, 1971) pp. 1-10.
2 Kalmen Cohen and Jerry Pogue. "An Empirical Evaluation of Alternative Portfolio Selection Models" *Journal of Business*, (April, 1968) pp. 180-190.
3 Edwin J. Elton and Martin J. Gruber. "Estimating the Dependence Structure of Share Prices" *Journal of Finance* (December, 1973) pp. 1203-1232.
4 Edwin J. Elton, Martin J. Gruber and Manfred W. Padberg. "Simple Criteria for Optimal Portfolio Selection" *Journal of Finance* (December, 1976) pp. 1341-1357.
5 Edwin J. Elton, Martin J. Gruber and Manfred W. Padberg. "Simple Criteria for Optimal Portfolio Selection—with Upper Bounds" *Operations Research.*
6 Edwin J. Elton, Martin J. Gruber and Manfred W. Padberg. "Simple Criteria for Optimal Portfolio Selection—Multi-Group Models *Journal of Financial and Quantitative Analysis.*
7 James L. Farrel. "Analyzing Covariation of Returns to Determine Homogeneous Stock Groupings" *The Journal of Business* (April 1974), pp. 186-207.
8 Lawrence Fisher and J. Kamin. "Good Betas and Bad Betas", Paper for the Seminar on the Analysis of Security Prices, University of Chicago, November, 1971.
9 B. King. "Market and Industry factors in Stock Price Behavior" *Journal of Business* (Jan, 1966), pp. 139-190.
10 Robert C. Klenkosky and John D. Martin. "The Adjustment of Beta Forecasts", *Journal of Finance*, (September 1975), pp. 1123-1128.
11 O. Vasicek. "A Note on Using Cross-Sectional Information in Bayesian Estimation of Security Betas", *Journal of Finance*, 28 (December 1973), 1233-1239.