INTERNATIONAL JOURNAL OF BUSINESS, 5(2), 2000 ISSN:1083-4346

Variation and Covariation between Market Timing and Selectivity: an Alternative to Traditional Meta-analysis

Sydney D. Howell[a] and Manuel J. Rocha Armada[b]
[a] Manchester Business School, Manchester, England
[b] School of Economics and Management, University of Minho, Braga, Portugal

ABSTRACT

In traditional meta-analysis of variance, using either Henriksson-Merton coefficients or Pfleiderer & Bhattacharya coefficients, both timing and selectivity show variation. Both models give similar estimates of the true selectivity variance. A new method of analysis exploits the fact (earlier ignored) that the selectivity and timing estimates are negatively correlated, due to measurement error and to correlated independents. This permits an estimate of the measurement error, and hence of the true variation. The new method implies that both the selectivity and the timing variance may not differ significantly from zero.

JEL classification: *G10, G11, G19*

Keywords: *Market timing; Selectivity; Performance evaluation; Variance and covariance analysis: meta-analysis*

Copyright©2000 by SMC Premier Holdings, Inc. All rights of reproduction in any form reserved.

I. INTRODUCTION

Correlation between estimates of timing and selectivity, and the true variability of timing and selectivity, have attracted interest recently. The concepts are, we argue, related. Coggin [5] show that selectivity and timing are negatively correlated for several categories of US fund managers, using both the Treynor [21] model and the Pfleiderer [17] model. Coggin [5] used the latter model in a form due to Coggin [6], that permits timing ability to be estimated as negative, a possibility rejected by earlier researchers. Coggin [6] show that negative correlation of Pfleiderer [17] estimates in the new form is equivalent to the positive correlation found when Lee [14] used the Pfleiderer [17] model in its original form. Coggin [5] argue that "the observed negative correlation in our data is largely an artifact of negatively correlated sampling errors for our two estimates", and that "much work needs to be done" on both the artefactual and substantive nature of these and other results. The Coggin [5] study, like many others, suggests that selectivity is on average positive, but timing ability is negative, in US markets.

Coggin [5] tried to estimate the true underlying variation of selectivity and timing by performing traditional meta-analysis on estimates from the revised Pfleiderer [17] model. Traditional meta-analysis compares the variation of parameter estimates between different studies, with the weighted average of their calculated confidence intervals within each individual study. The weighted intra-study confidence interval is taken to represent measurement error. Variation of parameter estimates between different studies that falls outside this confidence interval is taken to be real variation of the parameter, although Coggin [6] note that alternative explanations include "other unaccounted for artifacts". Coggin [5] and Coggin [6] conclude that there is significant real variation between managers, in both selectivity and timing ability, as measured by the Pfleiderer [17] and Treynor [21] models, so that "the best investors produced substantial risk adjusted excess returns". We show evidence that may weaken this conclusion.

The traditional meta-analysis technique adjusts carefully for covariation between different estimates of the *same* parameter, but it ignores the known covariation between the estimates of *different* parameters. This procedure risks discarding statistical information, and it also overstates the degrees of freedom available for estimating or testing each individual parameter. This paper argues that if negative covariation between the estimates of selectivity and timing is truly caused by measurement error, as Coggin [5] suggest, then the size of negative correlation must contain information about the size of the measurement error. In turn the size of the measurement error holds information about the "true" variation of the parameters.

II. OBJECTIVES OF THIS PAPER

This paper tries to construct an alternative method to traditional meta-analysis. The new method exploits the fact that measurement error can cause both variation and covariation of the selectivity and timing estimates. Hence the "true" variances can be inferred from the pattern of variance and covariance of the coefficients. Unlike traditional meta-analysis, the new approach can estimate true parameter variation from three different types of data, namely inter-fund variance-covariances only, intra-fund variance-covariances only, and the traditional method of comparing these both levels.

We compare traditional meta-analysis and the new method of analysis on estimates of selectivity and timing from the Henriksson [9] model (henceforward the Henriksson-Merton model). In this way we partly replicate and partly extend Coggin [5]'s traditional meta-analysis, which they applied to the Pfleiderer [17] model. We also replicate earlier work on the Treynor [21] model, the Henriksson [9] model and the original form of the Pfleiderer [17] model. We use a non US data set, which further extends the scope of previous studies, and we show that the Treynor [21], Pfleiderer [17] and Henriksson [9] models all behave similarly on US data and on UK data.

The sequence of the paper is as follows: Section III describes our replication of earlier work on the Treynor [21], Pfleiderer [17] and Henriksson [9] models, and our extension of traditional meta-analysis to the Henriksson [9] model. Section IV describes the negative correlation of selectivity and timing in our Henriksson [9] estimates. Section V presents a theoretical method for exploiting the variation and covariation of parameter estimates to estimate true parameter variation, in the presence of measurement error and correlated regressors. Section

VI describes an empirical application. Section VII is a summary and discussion.

III. ESTIMATES OF THE TREYNOR-MAZUY MODEL [21], THE HENRIKSSON-MERTON MODEL [9] AND THE PFLEIDERER-BHATTACHARYA MODEL [17] IN THE UK STOCK MARKET, PLUS META-ANALYSIS OF THE HENRIKSSON-MERTON PARAMETERS

The general aim of this section is to show how closely US and UK data sets agree, and to detect some internal evidence against the findings of traditional meta-analysis on Henriksson [9] data. For three different models (Henriksson [9], Pfleiderer [17] and Treynor [21]), our UK data set gave similar estimates to one or more US data sets, and the differences between the parameter estimates from US and UK data sets are similar to those between different US data sets.

We replicated Lee [14] by applying the original form of the Pfleiderer [17] model to the monthly returns of 141 UK unit trust funds, over a 144 month period from 1978 to 1990. We calculated both the GLS and OLS coefficients. The positive correlation between α_p (selectivity) and ρ (timing), was 0.486 for GLS coefficients (0.504 for OLS). The former compares closely with the figure of 0.47 reported by Lee [14] using GLS coefficients from US data. Mean levels of selectivity and timing, for our heteroscedasticity corrected estimates, were respectively -0.0001 for selectivity (standard deviation 0.00239) and 0.0687 for timing (standard deviation 0.0557). These estimates are of the original Pfleiderer [17] model, which forces all timing to be positive. Lee [14] corresponding mean estimates were 0.0008 for selectivity and 0.1231 for timing, so the differences are not significant.

Coggin [6] reanalysed the Lee [14] data, by allowing the Pfleiderer [17] model to estimate negative timing, and found after this adjustment that the same data implied negative correlation between selectivity and timing, estimated at -0.62 for the heteroscedasticity corrected coefficients, and -0.64 for the uncorrected coefficients. We have not replicated this adjustment for UK data, but we suspect that the UK correlation would be as close to the US correlation after adjustment as it was before. Coggin [5] applied the adjusted Pfleiderer [17] model to US pension fund data, and found a correlation of -0.44 between selectivity and timing. The above findings, and others, are summarized at the end of the paper in Table VI.

We calculated the Treynor [21] model for UK data, with heteroscedasticity correction, and found mean coefficients for selectivity of -0.0001 (standard deviation 0.00239) and for timing of -0.26 (standard deviation 0.479). These are broadly similar to US estimates by Coggin [5] of mean selectivity 0.000422 (SD 0.00265) and mean timing -0.279925 (SD 0.635032). Our correlation between Treynor [21] estimates of selectivity and timing was -0.57, whilst Coggin [5] found correlation between the Treynor [21] coefficients in the neighbourhood -0.45, using various market proxies and measures of correlation. These results suggest that our UK data set behaves much like the US data sets of Coggin [5], Coggin [6] and Lee [14].

We also replicated the US studies of the basic Henriksson [9] model [e.g. Chang [4], Jaganathan [11], Connor [7]]. Our UK results (discussed in more detail later) were again broadly similar to US results.

Having confirmed the comparability of US and UK data, we began to break new ground, by extending the Coggin [5] meta-analysis method to the parameter estimates of the Henriksson [9] model. Coggin [5] meta-analysed only two parameters, namely selectivity and timing, but we show in Sections II and III below that all three of the regression parameters are correlated, both theoretically and empirically. We therefore also meta-analysed the third parameter of the Henriksson [9] model, the upmarket beta.

Our traditional meta-analysis of the variance of Henriksson [9] selectivity gave similar results to the traditional meta-analysis of Pfleiderer [17] selectivity by Coggin [5]. We estimated the true variance for Henriksson [9] selectivity at 0.000002 (Coggin [5] estimated 0.000003 for Pfleiderer [17] selectivity: see Table VI). The similarity is interesting, since the gross inter-fund variation of selectivity is far larger in the Henriksson [9] model than in the Pfleiderer [17] (larger measurement error). Our estimate of the true Henriksson [9] selectivity variance did not differ significantly from a null hypothesis that it was zero, but equally it did not deviate significantly from a null hypothesis that it was equal to the Coggin [5] estimate.

The specifications of timing in the Henriksson [9] and Pfleiderer [17] models are incompatible, so the absolute values and variances of their timing coefficients cannot be compared. In our meta-analysis the true variation of Henriksson [9] timing was estimated to be significant, and was about 34% of its observed variance (Coggin [5] estimate for Pfleiderer [17] timing: 26.7%: of its observed variance). This implies that the Henriksson [9] timing estimate may be of higher statistical precision than the Pfleiderer [17] timing estimate, which seems unlikely.

We also meta-analysed the variance of the upmarket beta in the Henriksson [9] model. We estimated its true variance at 0.00868, which was significant, and a large percentage (57%) of its gross interfund variance. This also seemed implausible. Firstly, one might have expected well diversified and prudently managed funds to have rather similar betas, whereas this estimate implies that one fund in 20 had an average upmarket beta, over the entire 141 month period, either below 0.626 or above 0.995. Secondly, because there is high empirical correlation between all three Henriksson [9] regression parameters (selectivity, upmarket beta and timing) one would expect that if measurement error is high in any of them (as the large negative correlation of selectivity and timing suggests) it should be high in all.

These facts motivate a cautious approach to traditional meta-analysis of the Henriksson [9] model, and a more careful analysis of the relationship between measurement error, covariation and true variation for all three parameters in that model. We first consider the severe negative correlation between Henriksson [9] selectivity and timing.

IV. NEGATIVE CORRELATION BETWEEN SELECTIVITY AND TIMING IN THE HENRIKSSON-MERTON MODEL

Negative correlation between estimated timing and selectivity in the Henriksson [9] model has been reported by Henriksson [8], Chang [4], Jagannathan [11], Armada [2] and Connor [7]. Henriksson [8] and Jagannathan [11] found negative correlation for randomly chosen portfolios. The correlation is in the region of -0.86 for UK data (exact US correlations have not been published, to our knowledge). See Table 1 and Figure 1. Selectivity is identified in Table 1 as α, and timing as β_2, in reference to equation (1) below.

Table 1
Correlations of estimates of parameters for the Henriksson-Merton model. Sample: estimates of α, β_1 and β_2 from 141 models fitted to a cross-section of mutual funds. Dependent variable for each model: monthly returns 1978/90
(Individuals funds' values for α and β_2 are plotted in Figure 1).

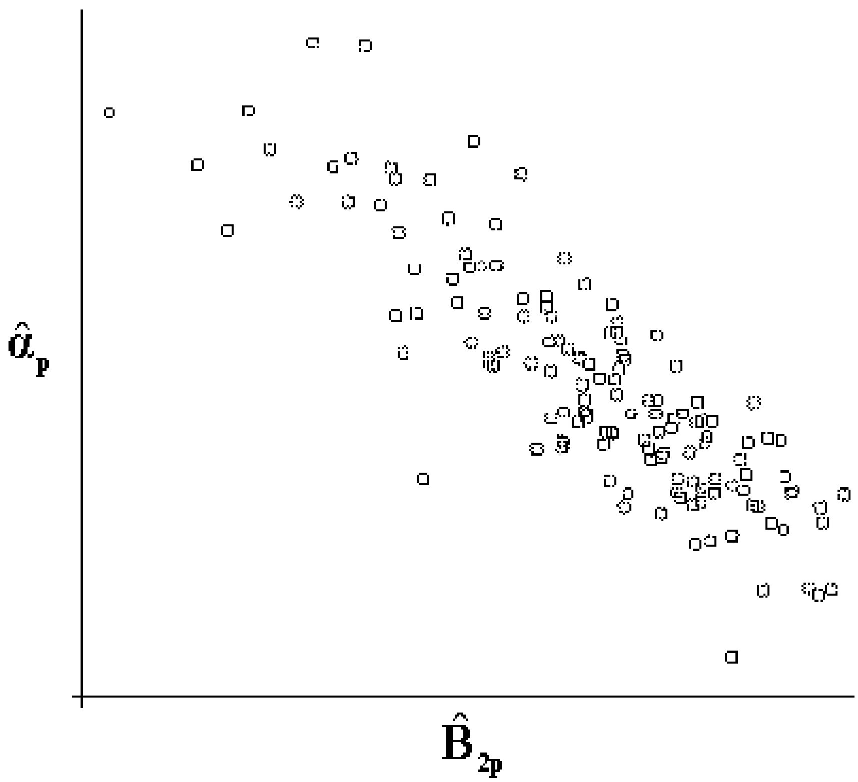
	α	β_1	β_2
α	1	-0.7201	-0.8571
β_1	-0.7201	1	0.6497
β_2	-0.8571	0.6497	1

Similar correlations seem to exist in US data and Japanese data. The proportion of funds that had at the same time significantly positive selectivity and significantly negative timing (or the opposite), which Chang [4] found in US data, is similar to the proportion found by Armada [2] in UK data. This suggests that negative correlations are similar in both markets[1].

In Table 1 (and Table 2 below), the correlations which Henriksson [8] and Coggin [5] would treat as interesting are the strong negative ones between estimated α and β_2 (selectivity ability and timing ability). However there are equally robust correlations between β_1 and β_2, and between α and β_1. These also require explanation, and, we argue, can be exploited in estimation.

Henriksson [8] states without analysis that measurement error "is not likely to be the entire explanation" for negative correlation, and he accordingly casts doubts on the CAPM. In contrast Coggin [5] quote from Hunter [10] that "*correlation between the estimates for selectivity and timing will necessarily be negative....because the sampling errors for the two estimates are negatively correlated. The magnitude of the correlation between the two estimates is the same as the magnitude of the negative correlation between the two sampling errors*." Coggin [5]'s statement is not quite technically correct, since it is valid for covariance rather than correlation, and a given covariance can lead to various levels of correlation, as discussed below.

Figure 1
Scatterplot of $\hat{\alpha}_P$ vs. $\hat{\mathrm{B}}_{2p}$ for 141 funds:
monthly returns: Feb. 78 – Feb. 90



Equation (1) is due to Merton [16] and Henriksson [9]:

$$Z_p(t) = \alpha_p + \beta_{1p}X(t) + \beta_{2p}Y(t) + \varepsilon_p(t) \quad (1)$$

where: $$Y(t) = MAX[0, -X(t)]$$

In this equation $Z_p(t)$ is the excess return (compared to the risk free asset) earned by fund p in time t, α_p is a return component to fund p which is independent of market movements, and X(t) is the excess of the market return over the risk free return in time t. The error term $\varepsilon_p(t)$ is heteroscedastic, which the estimation should in theory allow for. We found like Lee [14] that the correlation between the three coefficients in equation (1) is not sensitive to whether the estimates are OLS or heteroscedasticity adjusted.

Henriksson [9] interpret α_p as a measure of selectivity skill, and β_{2p} as a measure of timing skill, since Y(t) can be related to the return from a put option on the market, taking the value of zero when the market rises, and -X(t) when the market return falls below that of the risk free asset (i.e. when $X(t) < 0$). This assumes that the manager makes a timing move in response to the expected sign of a downward market movement, irrespective of its expected magnitude. The timing move can be to move out of the risky market, or to purchase a put on the market. The cost of this transaction is not theoretically allowed for in (1), but it is implicitly present in empirical estimates of (1), since it reduces the fund's overall assets, and hence its return.

Some assumed restriction on the investor's timing behaviour, not necessarily as severe as this one, seems essential if we are to estimate timing ability from portfolio return data, according to Admati [1], and Lehmann [15]. Our own results below, although they are noisy (or perhaps even because they are so noisy), appear to reinforce the doubts which Pfleiderer [17] raised on theoretical grounds, and which Connor [7] developed on empirical grounds, that the Henriksson [9] model may not be very efficient at detecting timing.

In the spirit of Coggin [5] we call the interfund comparisons of Table 1 and Figure 1 "meta-data". OLS estimates are displayed, since GLS estimates made little difference to the correlations (compare Tables 4a and 5a). We reasoned that if negative parameter correlation applies over cross sections of funds, it should also apply for each individual fund (the logic of traditional meta-analysis in reverse). If so, the calculated correlations between coefficients α and β_2 in (1), for each of the 141 individual funds in the sample, should also be negative.

Table 2 confirms this, to a degree which briefly startled us. The estimated OLS coefficient correlations for any one fund, as shown in this table of "micro" or intra-fund estimates, are identical for all 141 individual funds, to the highest accuracy available. The correlations change slightly between sampling periods, but are identical for all funds in a given period. Table 2's micro correlation is, as expected, similar to the meta-correlation in Table 1 across all 141 funds. However for α and β_2, the correlation in the meta-data is stronger than in the micro data. Traditional meta-analysis might interpret a larger variance and covariance in the meta data to mean real variation, but an alternative interpretation is that the meta-data are affected by measurement errors not foreseen at the micro level, in such a way that most of the extra variance is pure noise, and hence adds both variance and negative covariance to interfund estimates of selectivity and timing.

The reason that Table 2 is uniform across all funds is that the theoretical correlation matrix of OLS regression coefficients is based only on the correlations of the independent variables (including the constant vector of ones, when moments are taken about the origin). <u>These independents are identical for all the funds in a given time sample, thus yielding identical micro-estimates of the correlations of the regression coefficients</u>. This is a salutary reminder that large cross sections of funds may <u>not</u>, in important senses, yield statistically independent variation. GLS estimates of coefficient correlation do not show the perfectly uniform confidence intervals of OLS estimates, but they are very similar on average. They also suffer from additional noise, and from a loss of degrees of freedom, both arising from the estimation of the heteroscedasticity correction.

The mean levels of Henriksson [9] selectivity and timing which we estimated for the total period, using heteroscedasticity corrected estimates, were for selectivity, 0.0006 (standard deviation 0.0038), and for timing -0.077 (standard deviation 0.163). Comparable estimates without heteroscedasticity correction were, for selectivity 0.0008 (standard deviation 0.0036), and for timing -0.091 (standard deviation 0.153). The coefficients conform to the pattern of positive selectivity and negative timing found in many US studies. The heteroscedasticity correction tended on average, as theory predicts, to reduce the coefficients slightly, and to increase their variances. We next consider possible explanations for the correlations.

Table 2
Expected correlations of parameter estimates for the Henriksson-Merton model. Based on the standard joint confidence intervals for the regression parameters, for each of the 141 models used in Table 1 and Figure 1.

	α	β_1	β_2
α	1	-0.7203	-0.7704
β_1	-0.7203	1	0.8709
β_2	-0.7704	0.8709	1

Note: Each of the141 OLS models gives the identical correlation matrix shown below, for the overall time period 1978-90.

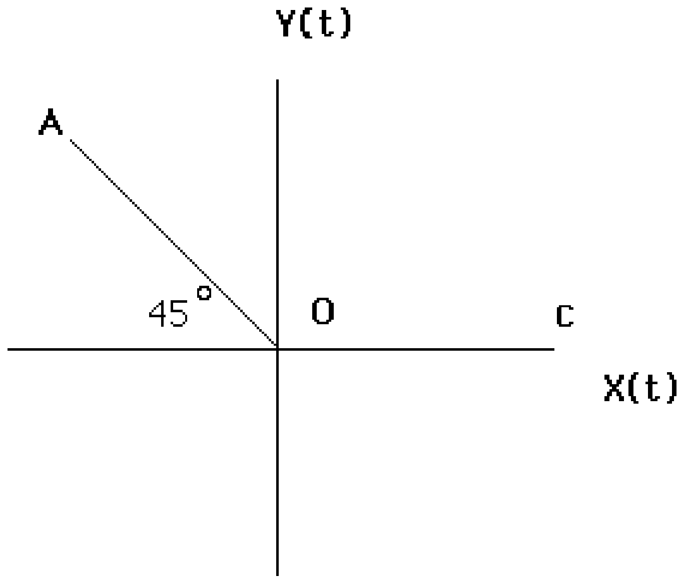
V. THEORY

A. Correlation between the two beta parameter estimates, due to correlation of the independent variables

Figure 2 plots the definitions of the two independent variables in (1), namely Y(t) and X(t). It shows definitional negative correlation between these variables, which causes positive correlation between their estimated coefficients β_1 and β_2.

Figure 2

The definitions of the two independent variables in (1): Y(t) and X(t)



If <u>all</u> observed X(t) are negative (line segment AO) there is perfect negative correlation between Y(t) and X(t). If <u>all</u> observed X(t) are positive (segment OC) there is zero correlation between Y(t) and X(t), since Y(t) is identically zero. But in general both positive and negative X(t) (positive and zero Y(t)) are observed, so that the mean value of Y(t) is positive. But if Y(t)'s mean is positive, even zero values of Y(t) (when X(t) is positive) become negative deviations from that mean, and contribute elements of negative covariance between X(t) and Y(t).

The resulting negative correlation between X(t) and Y(t) depends on the form of the distribution of the market return, and its location with respect to the risk free return. The macro estimate of correlation in Table 1, which is close to -0.86, is not very different from the cosine of angle OAC (-0.84) which is the expected correlation between X(t) and Y(t) for a uniform distribution of returns centred on the origin.

Alternatives to the Henriksson [9] model, such as the Treynor [21] and Pfleiderer [17] models, are also affected by correlation of the independents, but less strongly. For example when the market return is positive (as may be expected on average) it is positively correlated with its own square. We therefore expect the coefficients for these two variables in the Treynor [21] model, which is the first stage of the Pfleiderer [17] model, to be slightly negatively correlated. The absolute size of the correlation must in general be smaller than in the Henriksson [9] model of Figure 2, since the correlation takes opposite signs for positive and negative X(t), and is less than one in either case.

It would therefore be interesting to estimate all three models in a severe bear period, when average market return is negative throughout, and hence *negatively* correlated with its own square. We might expect a changed pattern of correlation between the parameters in all three models.

Negative correlation between the two independent variables can therefore explain the positive correlation between their estimated beta coefficients in (1) as in Table 1. We now consider the negative correlations between both β's and the constant α.

B. Negative correlation between α and β estimates

Negative correlation between regression estimates of α and both β's in (1) can be removed by re-expressing the independents in deviation form, where the underlined terms are means:

$$Z_p(t) = \alpha^*{}_p + \beta_{1p}[X(t) - \underline{X}(t)] + \beta_{2p}[Y(t) - \underline{Y}(t)] + \varepsilon_p(t) \tag{2}$$

By multiplying out and collecting terms we see a relation between (1) and (2):

$$\alpha_p = \alpha^*{}_p - \beta_{1p}\underline{X}(t) - \beta_{2p}\underline{Y}(t) \tag{3}$$

α^*_p and α_p are in general identical only if X(t) and Y(t) have means of zero (or of suitable value and opposite sign) which makes (1) identical to (2), but CAPM considerations require both the means to be positive in general. Although the least squares estimate of α^*_p in (2) is not correlated with estimates of β_{1p} and β_{2p}, α^*_p merely measures fund p's total average excess return over the risk free asset. Parameters of interest, such as selectivity ability α_p, are found by substitution using (3), which introduces negative correlation.

Hence the regression specification enforces negative correlation between empirical estimates of α and both β parameters, but gives no information on any intrinsic correlations of the parameters themselves. In fact, the regression model assumes that all parameters are fixed, so the question of variation and covariation between them cannot even arise. We try to relax this assumption in the next subsection, which models the relationship between true parameter variance, measurement error and correlation of the parameter estimates.

C. Small sample estimation behavior of identity relationships resembling the Henriksson-Merton model

Assume that investor performance can be defined identically as the sum of a selectivity component and a market timing component, both being normally distributed with constant variance. We treat an accurately observed total return T (analogous to Z_p in (1)) as the sum of two independent normally distributed random variates a and b (analogous respectively to the selectivity component α and the timing component $\beta_{2p}Y(t)$ in (1); we temporarily ignore $\beta_{1p}Y(t)$). All variables are in deviation form. Suppose some estimation process that accurately follows the random variation of b. The selectivity component a of each observation of return is estimated by subtraction as $\hat{a} = T - b$ (equation (6b) below). If a and b have variances of A and B respectively, and zero covariance, it is elementary that the estimate of a is accurate, and the estimates $\hat{a}$ and b are uncorrelated, as follows:

$$T = a + b \tag{4}$$

Derive the variance-covariance matrix for T, a and b respectively, by noting that for vector $X \sim N(\mu, \Sigma)$, its linear function QX is distributed $N(Q\mu, Q\Sigma Q')$, where X, Σ and Q respectively are (nx1), (nxn) and (pxn) (e.g. Press [18], p. 63). Here n is 2, p is 3, Σ is (2x2) and:

$$\Sigma = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \tag{4a}$$

We define the three rows of Q respectively from (4) and the identities a = a and b = b. Hence the variance-covariance of T, a and b is:

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} A+B & A & B \\ A & A & 0 \\ B & 0 & B \end{bmatrix} = Q\Sigma Q' \tag{5}$$

To find the variance-covariance matrix between estimator â for a and the accurately observed b, we substitute expression (5) (3x3) for Σ and use (6a) and the identity b = b respectively as the rows of a new Q (2x3). Equation (6b) shows that each resulting estimated observation â of random variable a has the correct variance A and is uncorrelated with b:

$$\hat{a} = T - b \tag{6a}$$

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} A+B & A & B \\ A & A & 0 \\ B & 0 & B \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \tag{6b}$$

Equation (1) converges in probability to an accounting identity resembling (4), in which the fixed parameter β_2 is observed without error. However in finite samples of (4), as both Armada [3] and Hunter [10] point out, if b is estimated with error, whilst total performance T is observed with or without error, each error in the estimate of b is on average matched by an equal and opposite error in the estimate of a, contributing a component of negative covariance between estimates of a and b. We can show that the negative covariance does not imply a unique level of negative correlation between the estimates, as Hunter [10] and Coggin [5] have suggested.

Define an error term e, having mean zero and variance E, uncorrelated with all other variables, which distorts observations of b. The estimator of b becomes $\hat{b}$ such that:

$$\hat{b} = b + e \tag{7}$$

And a redefined estimator â of a is again obtained by subtraction:

$$\hat{a} = T - \hat{b} \tag{8}$$

The variance-covariance matrix of â and $\hat{b}$, respectively, is defined by taking the rows of Q (2x3) to be the coefficients of these variables in (7) and (8):

$$\begin{bmatrix} 1 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} A+B & A & B & 0 \\ A & A & 0 & 0 \\ B & 0 & B & 0 \\ 0 & 0 & 0 & E \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ -1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} A+E & -E \\ -E & B+E \end{bmatrix} \tag{9}$$

Equation (9) shows that there is negative covariance between â and $\hat{b}$ [not, as Coggin [5] state, negative correlation] equal in magnitude to E, the magnitude of the error variance in estimating b. The correlation between â and $\hat{b}$ depends on the relative sizes of A, B and E. It approaches -1 if E approaches infinity, or if both A and B tend to zero for a non zero E. Such covariation seems unavoidable in finite samples. Equation (9) implies that there is (asymptotically) a relationship between the observed variances of â and $\hat{b}$ (respectively A+E and B+E), the error variance E, the covariance E and the true variances A and B. We try to exploit this relationship later.

We first attempt a richer approximation to the structure of (1). Assume that total investment return T is made up of three elements, the selectivity component a, the portfolio upmarket beta component b_1 [analogous to $\beta_{1p}X(t)$ in equation (1)] and a timing component b_2 [analogous to $\beta_{2p}Y(t)$ in (1)], so that:

$$T = a + b_1 + b_2 \tag{10a}$$

The three return components are assumed to be uncorrelated[2] and to have variances respectively A, B_1, and B_2. These yield the variance-covariance matrix below for the total return T and its three components a, b_1 and b_2 respectively:

$$\begin{bmatrix} A+B_1+B_2 & A & B_1 & B_2 \\ A & A & 0 & 0 \\ B_1 & 0 & B_1 & 0 \\ B_2 & 0 & 0 & B_2 \end{bmatrix} \tag{10b}$$

In order to model measurement error, we define error terms e_1 and e_2, having variances E_1 and E_2 respectively, each being uncorrelated with all other variables[3]. These error terms are assumed to affect the performance of estimators of b_1 and b_2 respectively. If:

$$T = a + b_1 + b_2 \tag{11a}$$

Let:

$$\hat{b}_1 = b_1 + e_1 \tag{11b}$$

So that, from previous assumptions, the variance of $\hat{b}_1$ is $B_1 + E_1$. We assume that some estimator is available for $\hat{b}_2$. We ensure that its performance models the known empirical covariance between regression estimates of β_{1p} and β_{2p}, due to correlation of the independent variables (as in Figure 2) by specifying as follows (other specifications are possible):

$$\hat{b}_2 = b_2 + k\hat{b}_1 + e_2 \tag{12}$$

The variance of this estimate is $B_2 + E_2 + k^2(B_1 + E_1)$. Since we are modelling the performance of <u>estimates</u>, the coefficient k is expected to be positive, because the known negative correlation between the <u>variables</u> X(t) and Y(t) in (1) leads to positive correlation between their coefficients β_1 and β_2 .These coefficients are modelled, indirectly at this stage, by $\hat{b}_1$ and $\hat{b}_2$.[4]

We redefine $\hat{a}$ as the residual after subtracting $\hat{b}_1$ and $\hat{b}_2$ from T:

$$\hat{a} = T - \hat{b}_1 - \hat{b}_2 \tag{13}$$

The variance-covariance matrix of the three estimators, in the sequence $\hat{b}_1$ and $\hat{b}_2$ and $\hat{a}$ respectively (still taking T to be measured without error), is:

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & k & 1 & k & 1 \\ 1 & 0 & -(k+1) & -1 & -(k+1) & -1 \end{bmatrix} \cdot \begin{bmatrix} A+B_1+B_2 & A & B_1 & B_2 & 0 & 0 \\ A & A & 0 & 0 & 0 & 0 \\ B_1 & 0 & B_1 & 0 & 0 & 0 \\ B_2 & 0 & 0 & B_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & E_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & E_2 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & k & -(k+1) \\ 0 & 1 & -1 \\ 1 & k & -(k+1) \\ 0 & 1 & -1 \end{bmatrix} = \tag{14}$$

$$= \begin{bmatrix} B_1+E_1 & k(B_1+E_1) & -k(B_1+E_1)-E_1 \\ k(B_1+E_1) & k^2(B_1+E_1)+B_2+E_2 & -k^2(B_1+E_1)-kE_1-E_2 \\ -k(B_1+E_1)-E_1 & -k^2(B_1+E_1)-kE_1-E_2 & A+k^2(B_1+E_1)+(1+2k)E_1+E_2 \end{bmatrix}$$

Relation (14) suggests important asymptotic relationships between the various variances, covariances and correlations in the Henriksson [9] type of model. Not all combinations of true parameter variation, covariation and measurement error are (asymptotically) possible. Traditional meta-analysis of the Henriksson [9] model ignores these relationships, which seems unwise. It may also be unwise that traditional meta-analysis of the Treynor [21] and Pfleiderer [17] models ignores any analogous relationships that may affect them.

The present model is however only illustrative. It ignores several potentially important refinements such as heteroscedasticity, and the non-normal distribution of k. Perhaps more seriously, it omits to model the fourth parameter of (1), the error term. We assume in (11a) that true performance is observed without error and is completely determined by fluctuations in true selectivity, true upmarket beta and true timing, but with no further random fluctuation. This is potentially unrealistic, but to allow for it would complicate the estimation as discussed later.

The notation of (14) can be simplified as follows[5]:

Defining the gross observed variances of $\hat{b}_1$, $\hat{b}_2$ and $\hat{a}$ as X, Y and Z:

$$X = B_1 + E_1 \tag{15}$$

$$Y = k^2 X + B_2 + E_2 \tag{16}$$

$$Z = A + Y + (1 + 2k)E_1 - B_2 \tag{17}$$

The variance-covariance matrix (14), in the revised row/column sequence $\hat{a}$, $\hat{b}_1$ and $\hat{b}_2$, can be expressed as:

$$\begin{bmatrix} Z & -kX - E_1 & -Y + B_2 - kE_1 \\ -kX - E_1 & X & kX \\ -Y + B_2 - kE_1 & kX & Y \end{bmatrix} \tag{18}$$

Correlations are computed by dividing each covariance by the square root of the product of the respective individual variances:

$$Corr(\hat{b}_1, \hat{b}_2) = kX \Big/ \left[(XY)^{0.5}\right] \tag{19}$$

$$Corr(\hat{b}_1, \hat{a}) = -(kX + E_1) \Big/ \left[(XZ)^{0.5}\right] \tag{20}$$

$$Corr(\hat{b}_2, \hat{a}) = -(Y - B_2 + kE_1) \Big/ \left[(YZ)^{0.5}\right] \tag{21}$$

Only one covariance parameter is explicitly assumed in (14) namely k, whose existence, due to the correlation of independents, is well established. The rest of the highly sensitive covariance structure of (14) is due to the measurement errors E_1 and E_2.

This sensitivity is illustrated in Table 3.

Table 3
Parameter sets giving approximately "expected" values for intercoefficient correlations

	1	2	3	4	5	6
A	0.250	0.080	0.220	0.350	0.210	0.150
B_1	0.450	0.015	0.300	0.060	0.007	0.320
B_2	0.100	0.010	0.100	0.015	0.004	0.070
E_1	0.120	0.030	0.040	0.200	0.113	0.030
E_2	0.050	0.008	0.030	0.040	0.012	0.014
k	0.800	1.100	1.050	0.780	0.670	0.850

Intercoefficient correlations generated by the above

b_1,b_2	0.84	0.87	0.86	0.86	0.88	0.87
b_1,a	-0.77	-0.77	-0.79	-0.77	-0.76	-0.77
b_2,a	-0.72	-0.73	-0.73	-0.75	-0.73	-0.72

Note: The above results take all parameters at the scale of A (raw monthly excess return). The scaling effects of X(t) and Y(t) on parameters B_1 and B_2 are omitted.

In Table 3 we vary all six assumed parameters, in order to imitate the observed values of three intercoefficient correlations. Not surprisingly, since there are three surplus degrees of freedom, some very diverse solutions are possible. The correlations are sensitive to small variations in the assumed parameters. The sensitivity has important implications when we try later to estimate the parameters themselves from the observed correlations.

We tried the effect of reducing k, to see whether a lower and/or reversed covariation between the independent variables (as expected in the Treynor [21] and Pfleiderer [17] models) leads in general to smaller negative correlation between selectivity and timing. We found no simple relationship between k and the level of correlation. This, plus various structural differences between the Henriksson [9] and the other models, makes it hard to foresee the result of applying the present modelling approach to the Treynor [21] or Pfleiderer [17] models.

D. Estimation from asymptotic relationships

We wished to estimate from an empirical sample of (14) the six parameters within it, namely the hypothesised "true" parameter variances A, B_1 and B_2, [which are the central issue addressed by Coggin [5]] and the two hypothesised error variances E_1 and E_2, together with k.

Interestingly, there is an exact solution. The six unobservable parameters assumed in (14) generate a total of six observable parameters (namely the observed values of variances X, Y and Z, and their three intercorrelations). We can solve exactly for a point estimate of all six unobservable parameters, using simple substitution on the six observable values, as follows. Rearranging (19):

$$k = Corr\left(\hat{b}_1, \hat{b}_2\right)\left(XY^{0.5}\right)\Big/X \qquad (22)^6$$

Rearranging (20) and substituting from (22):

$$E_1 = -\left[Corr\left(\hat{b}_1, \hat{a}\right)\right]\left(XZ\right)^{0.5} - kX \qquad (23)$$

Rearranging (21) and substituting from (22) and (23)

$$B_2 = Corr\left(\hat{b}_2, \hat{a}\right)\left(YZ\right)^{0.5} + Y + kE_1 \qquad (24)$$

And by substitutions and rearrangements in (17), (16) and (15):

$$A = Z - Y - (1 + 2k)E_1 + B_2 \qquad (25)$$

$$E_2 = Y - B_2 - k^2 X \qquad (26)$$

$$B_1 = X - E_1 \qquad (27)$$

An interesting feature of this approach is that unlike traditional meta-analysis, it uses only a single variance-covariance matrix, so it can be solved either using inter-fund covariances alone, or intra-fund covariances alone. Comparison between the two estimates, essential in traditional meta-analysis, is not necessary, but is also possible.

The point estimates produced by equations (22) to (27) are sensitive to small variations in the inputs, so it is best to work at the highest available precision. Since the input parameters are derived from variances and covariances, it is theoretically desirable to use GLS estimates of them.

E. Testing problems

The estimates must, at best, have wide, highly skewed and jointly dependent distributions, which unfortunately are not convenient to treat. Their distribution depends ultimately on the joint distribution of the variance-covariance matrix of the regression estimates. A variance-covariance matrix of order p (under multivariate normality) is jointly distributed as Wishart (Σ, p, n), where the scale matrix Σ is the hypothetical true population covariance matrix, and n is the number of degrees of freedom, or fully independent observations of the covariance (see Press [18], p. 100 et seq.).

The Wishart distribution involves weighted sums of chi squared variates, which are processed by the present model in various ways. The Wishart distribution is not convenient to handle, and particular problems arise in the present case. For example it is not straightforward to test the joint hypothesis (either as a null or as an alternative) that all estimated variances are non negative, and that (for example) variances A and B_2 are both zero. A hill-climbing model is probably required. Other complications are that the expected Wishart scale matrix must depend on the particular estimator used for the Henriksson [9] model.

We do not pursue the formal testing project in this paper, but meanwhile as a descriptive statistic, we use the ratio of two variances, namely our estimate of (true variance plus noise variance) and our estimate of noise variance alone, for each parameter singly. The null hypothesis (of zero true variance) predicts a ratio of one. Coggin [5] use the same ratio to compare the equivalent outputs of traditional meta-analysis. In the context of their model, Coggin [5] multiply the variance ratio by the sample degrees of freedom (n-1) and treat the product as a χ^2 statistic with n-1 degrees of freedom. For descriptive purposes we perform a similar test, though its formal relevance in the present model is less clear-cut.

VI. AN EMPIRICAL APPLICATION

1. Method

We solved equations (22) to (27) using inputs derived from the regression coefficient variances of equation (1), which was fitted to the monthly returns of 141 UK unit trusts over subsets of a 144 month period, 1978-90. Our UK market proxy was the Financial Times All Share Index.

2. Scaling aspects

Equation (14) is expressed in terms of elements of monthly return, but the available input variances are those of the regression coefficients of (1) whose scale is different. For example the monthly return component b_1 due to the upmarket beta, used in (14), corresponds to $\beta_{1p}X(t)$ in (1). The variances of β_{1p} and $\beta_{1p}X(t)$ are several orders of magnitude apart, so the estimated variances of β_{1p} and β_{2p} must be rescaled in order to yield variances X and Y. Of course α is defined from the outset as a component of periodic return, so its variance as estimated in (1) need not be rescaled for use as Z in (22) to (27).

We rescaled the regression variances for β_{1p} and β_{2p} to the scale of variances of components of monthly return, by multiplying them respectively by $\underline{X}(t)^2$ and $\underline{Y}(t)^2$, using var(ax) = a^2var(x). We used the rescaled values as estimates of X and Y for equations (22) to (27). It is possible re-notate the entire model in terms of the variances and covariances of the original regression coefficients, but this is not essential. As an example, if the postulated "true" variances of the regression coefficients β_{1p} and β_{2p} are called B_1^* and B_2^* respectively, then (10b) becomes[7]:

$$\begin{bmatrix} A+\underline{X}+(t)^2 B_1{}^* + \underline{Y}(t)^2 B_2{}^* & A & \underline{X}(t)^2 B^*{}_1 & \underline{Y}(t)^2 B_2{}^* \\ A & A & 0 & 0 \\ \underline{X}(t)^2 B_1{}^* & 0 & \underline{X}(t)^2 B_1{}^* & 0 \\ \underline{X}(t)^2 B_2{}^* & 0 & 0 & \underline{Y}(t)^2 B_2{}^* \end{bmatrix} \tag{28}$$

The six input variables exist in two forms:
a) for the cross section of funds (meta-data), and
b) for all 141 funds individually (micro data).

a) the cross section of all 141 funds (see Table 1) yields one set of 6 output parameters, which estimate *inter alia*, the "true" inter-fund variation of timing, selectivity and upmarket beta. These estimates are in Tables IVa (OLS variances) and Table Va (GLS variances) for various time samples.

b) regression estimates of the six input variances and correlations exist for each individual fund, although in the OLS case the three correlations are identical for all funds (see Table 2) and the GLS correlations are also related but highly noisy (compare the resulting means and extremes of the output variables in Tables 4b and 5b). This yields 6 input and 6 output parameter estimates of intra-fund variation for each of the 141 funds, though highly dependent. The 141 micro estimates are exemplified and summarized in Table 4b (for OLS variances) and Table 5b (for GLS variances).

3. Scaling of results

Equations (22) to (27) estimate relationships between the variances of three elements of periodic return. However it is usual to discuss the "true" variances of the regression coefficients themselves, as estimated for (1) (rather than the variances of elements of return). Therefore the data in Tables 4a and 4b and 5a and 5b are reported at the scale of the regression coefficients of (1). This was done by reversing the earlier rescaling i.e. we divided the estimates of B_1 and B_2 from equations (22) to (27) by $\underline{X}(t)^2$ and $\underline{Y}(t)^2$ respectively, and similarly the estimates of E_1 and E_2.

For strict consistency with the notation of (28) the rescaled estimates of the true and noise variances of the regression coefficients could be labelled as B_1^* and B_2^* and E_1^* and E_2^*, but the table captions of Tables 4 and 5 omit this detail, since the final scale is pointed out in their titles.

4. Calculation aspects: OLS inputs versus GLS inputs

Unfortunately, although the GLS variances are asymptotically superior inputs to (14), they proved less stable than the OLS variances. The cause may be that GLS must estimate extra, and sensitive, parameters, from noisy data. The average estimated variance of a GLS coefficient was close to, but larger than, that of an OLS coefficient, as theory predicts (compare the first three rows between Tables 4a and 5a). However, in individual cases, the GLS variance was often much smaller or larger than the corresponding OLS variance. The OLS estimates, therefore, although slightly biased on average, probably have a much smaller total mean squared error. We therefore report the results using both OLS and GLS inputs.

5. OLS results: Tables 4a and 4b

Table 4a shows the "meta level" results from the inter-fund data, while Table 4b summarizes the 141 micro results of individual funds. For brevity, Table 4b shows only three individual funds in full detail, plus the average, maximum and minimum values over all 141 individual funds. Averages and extremes are reported for descriptive purposes, and may not be suitable for inference.

Tables 4a and 4b conflict sharply with traditional meta-analysis results. They suggest that on average there is little or no real difference between funds in timing ability (B_2 is zero), or in the upmarket beta (B_1 is zero). The OLS variances and covariances of the estimates can be explained on average as entirely due to the specified artifacts and to measurement error. There seems some case for true variation in selectivity (A) within or between funds, since A is a robustly large percentage of the estimated measurement error in both tables. This however is itself possibly an artifact, as we discuss below.

A descriptive variance ratio can be calculated, by adding one to the relevant fraction in the "ratios" section of each table. If the χ^2 test of Coggin [5] is assumed relevant, its value would be highly significant for

A in all periods and in both tables, and the value of B_2 in the second sub period of Table 4a would also be significant at 5%. On the more conservative assumptions of the f test, none of the true inter-fund variances in Table 4a is significant, and in Table 4b only A reaches significance, which measures intra-fund variation. We do not assume that either χ^2 or f is the exact distribution of this variance ratio.

Table 4a
Input variances calculated by OLS

Solutions of equations (22) to (27) for a cross section of 141 mutual funds are reported at the scale of the original regression coefficients, not of components of monthly return.

		Total Period 1978-90	Subperiod 1978-84	Subperiod 1984-90
Inputs (observed)	X	0.015360	0.026333	0.021162
	Y	0.023428	0.057630	0.028916
	Z	0.000013	0.000034	0.000017
	Corr $(\hat{b}_1, \hat{b}_2)$	0.6497	0.6786	0.6907
	Corr $(\hat{b}_1, \hat{a})$	-0.7201	-0.6817	-0.7545
	Corr $(\hat{b}_2, \hat{a})$	-0.8571	-0.8697	-0.7951
Scaling	X(t)	0.00844072	0.00910837	0.00777307
	Y(t)	0.01606381	0.01460923	0.01751838
Outputs (calculated)	A	0.000003	0.000010	0.000004
	B_1	0.000168	-0.002165	0.001732
	B_2	-0.000018	-0.007960	0.004210
	E_1	0.015192	0.028498	0.019429
	E_2	0.013557	0.039051	0.010911
	k	1.527038	1.610192	1.819647
Ratios	B_1/E_1	1.11%	-7.60%	8.92%
	B_2/E_2	-0.13%	-20.38%	38.58%

A/(Z-A)	28.00%	42.13%	29.22%

Table 4b

Coefficients calculated by OLS (inputs and outputs rounded)

Solutions of equations (22) to (27) for individual funds: summary of 141 UK Mutual Funds. Equations were solved at scale of monthly return, but the solutions are shown here at the scale of the regression coefficients.

		Total Period 1978-90					
		Three Sample Funds			Summary of 141 Funds		
		ABBG	BARUG	EFMGI	Average	Max	Min
Outputs (calculated)	A	0.000002	0.000002	0.000002	0.000005	0.000014	0.000001
	B_1	0.000000	0.000000	0.000000	0.000000	0.000006	-0.000005
	B_2	0.000000	0.000000	0.000000	0.000000	0.000000	-0.000002
	E_1	0.002428	0.003461	0.003297	0.006674	0.020608	0.001394
	E_2	0.001360	0.001938	0.001846	0.003739	0.011545	0.000781
	k	2.52391	2.52386	2.52389	2.52891	2.52393	2.52386
Ratios	B_1/E_1	-0.02%	-0.01%	0.00%	0.00%	0.08%	-0.09%
	B_2/E_2	-0.02%	-0.01%	-0.01%	-0.01%	0.01%	-0.03%
	A/(Z-A)	65.68%	65.68%	65.67%	65.67%	65.69%	65.66%

Note: Summary rows and columns generally relate to different individual funds: extremes of ratios need not arise from extremes of individual variables, and ratios of averages do not equal averages of ratios.

Among many striking uniformities in Table 4b, E_1 is a constant multiple of E_2. Equation (14) explained the empirical correlation of α and β_2 by postulating two uncorrelated error variances E_1 and E_2, so we were at first startled to find a further unexplained correlation arising in the empirical data, between E_1 and E_2 themselves. The explanation seems important.

Formally, the absolute variance-covariance matrix of the OLS regression coefficient estimates for fund p is simply the inverse of the variance-covariance matrix of the independent variables (which is identical for all the funds), multiplied by s_p^2, which is the standard error of estimate for fund p (see e.g. Wonnacott [22], p. 248). Hence the entire content of Table 4b possesses in theory only a single degree of freedom[8] reflecting the difference from fund to fund in the standard error s_p^2 of the regression fit of (1). Instead of 141 estimates of 6 covariance parameters, Table 4b reflects only 1 estimate of six parameters, varied by 141 estimates of one parameter. This fact has sobering implications for a large class of multi-fund comparisons.

6. The effect of the heteroscedasticity correction: Tables 5a and 5b

Tables 5a and 5b use heteroscedasticity-corrected estimates of the input variables. The results are disappointingly noisy. The GLS values of the input variances X, Y and Z are close on average to the OLS estimates, although slightly larger, as theory predicts (compare the input rows of Table 4a and Table 5a). But the resulting estimates of the output variables are much more scattered, especially for the individual funds, and over the shorter time periods. However when averaged over longer periods, and across all funds, the outputs of GLS are broadly similar to those of OLS.

Table 5a for the total period seems to show no true variation between managers in timing ability, and little or no true variation in upmarket beta. Both of these findings contradict traditional meta-analysis. Findings for the two subperiods are noisier, but the deviations of B_1 and B_2 from their null hypothesis mean of zero are broadly similar in direction to the OLS estimates in Table 4a, though larger in magnitude. Coggin [6] point out that if the true value of a variance such as B_1 or B_2 is zero, its value when estimated by subtraction within a finite total, as here, will be below zero with 50% probability, which is roughly what we find in both the GLS and OLS estimates (though at rather different scales). The absolute deviations of the GLS B_2 from zero are equal and opposite in the two sub periods.

Table 5a
Coefficients calculated with heteroscedasticity correction

Solutions of equations (22) to (27) for a cross section of 141 Mutual Funds. Solutions are at the scale of the original regression coefficients, not of components of monthly return.

		Total Period 1978-90	Subperiod 1978-84	Subperiod 1984-90	
Inputs (observed)	X	0.015640	0.026989	0.020672	
	Y	0.024068	0.057800	0.030606	
	Z	0.000014	0.000030	0.000016	
	Corr ($\hat{b}_1, \hat{b}_2$)	0.6553	0.6794	0.6764	$= \frac{kX}{(XY)^{0.5}}$
	Corr ($\hat{b}_1, \hat{a}$)	-0.7237	-0.6875	-0.7108	$= \frac{-(kX + E_1)}{(XZ)^{0.5}}$
	Corr ($\hat{b}_2, \hat{a}$)	-0.8691	-0.8679	-0.7675	$= \frac{Y - B_2 + kE_1}{(YZ)^{0.5}}$
Scaling	X(t)	0.008441	0.009108	0.007773	
	Y(t)	0.016064	0.014609	0.017518	
Outputs (calculated)	A	0.000003	0.000008	0.000004	
	B_1	0.000864	0.002111	0.006425	
	B_2	0.000122	-0.005007	0.005150	
	E_1	0.014775	0.024879	0.014246	
	E_2	0.014946	0.036127	0.011453	
	k	1.520514	1.594711	1.854901	
Ratios	B_1/E_1	5.85%	8.48%	45.10%	
	B_2/E_2	0.81%	-13.86%	44.97%	
	A/(Z-A)	24.47%	36.04%	34.84%	

The descriptive variance ratio reaches larger values in Table 5a than in 4a. This must be due to greater noise, which the χ^2 test ignores. If treated as χ^2 with (n-1) degrees of freedom, the variance ratio for A would be significant at 5% in every case, and the excursion of B_2 in the second subperiod would also be significant, as it was in the OLS estimates of Table 4a. The rounded estimates of A in Tables 5a and 5b respectively resemble the corresponding values in 4a and 4b (respectively 0.000003 and 0.000002). These estimates are not independent.

Table 5b

Variances calculated with heteroscedasticity correction (inputs and outputs rounded)

Solutions of equations (22) to (27) for individual funds: summary of 141 UK Mutual Funds. Solutions are at the scale of the original regression coefficients, not of components of monthly return.

		Total Period: 1978-90					
		Three Sample Funds			Summary of 141 Funds		
		ABBG	BARUG	EFMGI	Average	Max	Min
Outputs (calculated)	A	0.000002	0.000002	0.000002	0.000004	0.000014	0.000001
	B_1	0.001189	0.002249	0.000465	0.002910	0.036672	-0.004502
	B_2	-0.000037	0.000038	-0.000213	0.000000	0.008017	-0.001488
	E_1	0.001696	0.001980	0.003025	0.004180	0.018629	-0.017828
	E_2	0.001517	0.002200	0.001877	0.004038	0.012884	0.000734
	k	2.50111	2.54115	2.41112	2.58955	3.79439	1.80784
Ratios	B_1/E_1	70.12%	113.58%	15.36%	17.06%	1329.78%	-2824.34%
	B_2/E_2	-2.44%	1.74%	-1.14%	-3.66%	62.22%	-32.88%
	A/(Z-A)	65.80%	64.37%	70.38%	69.35%	92.46%	62.36%

Note: Summary rows and columns generally relate to different individual funds: extremes of ratios need not arise from extremes of individual variables, and ratios of averages do not equal averages of ratios.

Parameters A and k seem to be robustly positive in all the tables. Of course k is known to be an artifact, and we suspect that A is also. The reason is that if there is a random element in investment performance (net of variations of timing and selectivity ability), the model of (14) forces all its variance to be included in the selectivity variance A. Our robust estimate of A may therefore be in whole or in part an estimate of the random element of investment performance. It is not surprising that this estimate should be similar between mico and macro data, or even between the Henriksson [9] model and the Pfleiderer [17] model used by Coggin [5]. Unfortunately, if one tries to allow for the random variation by adding an explicit error term to (11a), the number of parameters grows to seven. These cannot be estimated from six variances and covariances, in the absence of additional information. In general, the presence of an extra parameter seems to reduce the estimates of the other variance parameters (see Appendix). This suggests that our present estimates of all true and noise variances may be biased upwards.

Subject to more analysis of the inference problems, a conservative interpretation is that our data set is compatible with a null hypothesis that the true variances of Selectivity and Timing are close to zero. The high noise level in the estimates means, however, that they are compatible with many other hypotheses. Indeed if our robust estimate of A is taken to be non-artifactual, our results actually go beyond the Coggin [5] interpretation, since they estimate a similar level of true variation in selectivity ability, but without any offsetting variation in timing. We suggest a further approach to this issue in Section V.

7. Comparison of the micro and macro estimates

A final step is is based on the null hypothesis of traditional meta-analysis, namely that the inter-fund variances and intra-fund variances should be similar. Large discrepancies between the two estimates would suggest (in this context) errors in the method. This can be tested by subtracting the estimated intra-fund variance in Table 4b or 5b from the estimated intra-fund variance in Table 4a or 5a, as in traditional meta-analysis.

The test shows no evidence of problems. For the true selectivity variance A, and for the true variance of the upmarket beta, B_1, the inter-fund variances, in Tables 4a and 5a, are actually smaller than the corresponding intra-fund estimates, in Tables 4b and 5b, though probably not significantly so. For the true timing variance B_2, the inter-fund variances are larger than the average of the intra-fund estimates, but both estimates are negligibly small on average, compared to the total observed variance Y of the timing parameter.

VII. SUMMARY AND DISCUSSION

Traditional meta-analysis gives similar results in the Henriksson [9] model and in the Pfleiderer [17] model, even if the Henriksson [9] is applied to UK monthly data and the Pfleiderer [17] to US monthly data. In both models the "true" variance of timing is estimated to be significant, and the "true" variance of selectivity is close to 0.0000025. The "true" selectivity variance seems statistically significant in the Pfleiderer [17] model, but not in the Henriksson [9] model. There are reasons for doubting these traditional meta-estimates, at least for the Henriksson [9] model.

a) Traditional meta-analysis ignores the known large covariation between selectivity and timing (and the upmarket beta), which is due to measurement error. This procedure discards useful information about the measurement error, and hence about true variation. It may estimate implausible joint states for the true variances, measurement errors and covariances of the three regression parameters.

b) Dependence is higher than generally recognized between the estimates for different funds in cross sections of funds. Degrees of freedom are also consumed by parameters not of direct theoretical interest.

c) Traditional meta-analysis gives an implausibly large estimate for the true inter-fund variance of the upmarket beta in the Henriksson [9] model (this has not been checked for the Treynor [21] or Pfleiderer [17] models).

d) Traditional meta-analysis does not remove parameter correlation. An individual manager who achieves a high estimate of "true" Henriksson [9] selectivity on a univariate meta-analysis test, will usually achieve a low estimate of "true" timing on the same test. So the manager's overall excess return in the sample period is far lower than traditional meta-analysis of timing or selectivity alone would suggest.

We have suggested a new type of analysis. This actively exploits the observed variation and covariation of all three Henriksson [9] parameters (selectivity, timing and upmarket beta) in order to estimate their true variation. The new approach, in a somewhat simplified and noisy implementation, contradicts traditional meta-analysis in several ways. It suggests that the true inter-fund variances of Henriksson [9] timing and selectivity may both be close to zero. It also suggests a lower, more plausible value for the inter-fund variation in the upmarket beta.

We have noticed that GLS estimates of variance can be very unstable in detail, and we have also shown that cross sectional studies can yield highly (even totally) dependent parameter estimates.

If, as we suspect, the true Henriksson [9] selectivity and timing variances are close to zero, it would not be true to conclude for the Henriksson [9] model, as Coggin [5] concluded from meta-analysis of the Pfleiderer [17] model that "the best managers produced substantial risk adjusted excess returns". However even if there exists some true selectivity variance, and even if this is not offset by poor timing, it remains an open question whether differences in selectivity within any one period persist over successive time periods.

1. Discussion

Why does our new analysis fail to detect true timing variation in the Henriksson [9] model? It may be that the Henriksson [9] model itself is not efficient at detecting timing. Pfleiderer [17] point out that the Henriksson [9] information specification is coarse, and Connor [7] offer relevant empirical evidence. It would be interesting to develop a version of our approach to make a comparable analysis for the Pfleiderer [17] model, though the problems seem formidable. This gap in our knowledge, and the many other gaps in Table 6, offer considerable scope for further theory and comparative work, as well as for replications in different time periods, in bull and bear markets etc..

Table 6
Summary of comparative results (All are GLS estimates)

	Lee [14], Coggin [6]	Coggin [5]	Armada [3]
Mean values of selectivity			
Selectivity Treynor [21]		0.000422	-0.0001
Selectivity Henriksson [9]			0.0006
Selectivity Pfleiderer [17] revised	0.000811	0.000339	
Mean values of timing (parameter levels are not comparable between models)			
Timing Treynor [21]		-0.279925	
Timing Henriksson [9]			-0.077
Timing Pfleiderer [17] revised	-0.084774	-0.046979	
Correlations of selectivity and timing			
Treynor [21]		-0.45	-0.57
Henriksson [9]			-0.87
Pfleiderer [17] original	0.47		0.486
Pfleiderer [17] revised	-0.62	-0.44	
Observed total meta variations of selectivity parameters (between funds)			
Selectivity Treynor [21]		0.000007	0.000006
Selectivity Henriksson [9]			1.33604E-05
Selectivity Pfleiderer [17] revised	0.000034	0.000007	
Estimated "true" variations of selectivity parameters (between funds)			
Meta true Variance of selectivity Treynor [21]		0.000004	
H-A true Variance of selectivity Treynor [21]			
Meta Variance of selectivity Henriksson [9]			0.000002
H-A Variance of selectivity Henriksson [9]			0.000003
Meta Variance of sel. Pfleiderer [17] revised	0.000022	0.000003	
Variations of timing parameters (between funds) Note: as the timing parameters are not comparable between Models, results are grouped per model			
Treynor [21] model			
Gross variation of timing Treynor [21]		0.403226	
Meta-analysis of true timing		0.345976	
H-A analysis of true timing			
Henriksson [9] model			
Gross variation of timing Henriksson[9]: (Y in 4a)			0.02342799
Meta-analysis of true timing			0.00795527
H-A analysis of true timing			-0.00001804
Revised Pfleiderer [17] model			
Gross variation of timing Pfleiderer [17]	0.038888	0.011027	
Meta-analysis of true timing	0.028776	0.002941	
H-A analysis of true timing			

The argument as to whether true variances in timing or selectivity exist in any sample, and truly persist over successive samples, can also be addressed by more conventional methods. Both traditional meta-analysis and our own method make a quantified estimate of the percentage of "true" variation, in the total observed variation of selectivity or timing. Under suitable assumptions, this predicts the "signal to noise" ratio in the selectivity or timing performance of the average manager. From this basis it is a problem in regression theory to predict what percentage of a given fractile of selectors or timers in one sample is expected to survive in the same fractile, in a second or later sample (such survival rates should be smaller in the noisy Henriksson [9] model than in the less noisy Pfleiderer [17] model).

Such a procedure could provide a test of whether persistently superior selectors and timers do exist, in the proportions that Coggin [5] suspect, as against a null hypothesis that true selectivity and timing are zero on average, and vary randomly, if at all, between successive time periods. We have made a preliminary survey of this type on our Henriksson [9] data, and found a correlation of approximately zero between the estimated selectivity or timing of each manager in two successive periods. This appears to support the null hypothesis against traditional meta-analysis, at least for the Henriksson [9] estimates, but the problems of interpretation are great.

An interesting feature of this approach is that it might be possible to prove that consistently successful stock selectors or timers exist in general, but hard to identify them as individuals. Coggin [5]'s statement that "the best equity pension fund managers delivered substantial risk adjusted excess returns" does not imply its converse, namely that the managers who delivered substantial risk adjusted excess returns in the Coggin [5] sample were the "best" managers (in long term expectation). If there is a hypothetical "true" best decile of selectors in long term expectation, the fact that there is a fairly high signal to noise ratio according to Coggin [5] or ourselves, implies that the "true long term" best decile of selectors have a fairly modest probability of appearing among the top decile of selectivity performers in any one sample, and a negligible probability of doing so in every sample. Interesting problems of sample design therefore arise, for testing any specific hypothesis on the true variances or persistences of selectivity and timing.

Leaving aside the problem of the variances of selectivity and timing for a moment, the problem of their means remains. Why does selectivity tend to be positive on average, and timing negative, for many models, and in more than one market, and for more than one type of manager?

Explanations could be substantive or artefactual. We suspect a bias by some researchers towards hypotheses that successful selectivity is substantive, and that apparently unsuccessful timing is an artifact. This is encouraged by the rumoured success of automatic trading systems (such success, if not correlated with market movements, will be estimated by the Henriksson [9] and Pfleiderer [17] models as selectivity). The fact that managers are willing to invest large sums in company analysis, in an otherwise almost efficient market, is further circumstantial evidence in favour of true selectivity. We know of little work on more pessimistic or artifactual hypotheses.

Ultimately, however, the means of selectivity and timing must be less interesting than their variances. Correctly defined selectivity and timing profits must sum (and also average) to zero across all players in all financial markets (including occasional players, and all market entrants and leavers during the sample period). Suitably measured selectivity and timing abilities should therefore also average to zero across all players, subject to estimation artifacts and sampling bias.

If a truly unbiased measurement method reports a non-zero mean for selectivity or timing, this might arise from sampling variation, or because the sample is biased towards investors who share advantages or disadvantages (shared bias should of course bias downwards the observed variances of the abilities). Samples of professional fund managers, and particularly from the largest funds, may well share economies of scale in collecting and processing information or transactions, or hiring superior managers, or implementing automatic trading systems. Some financial institutions might, in contrast, benefit more from scale economies in marketing and administration rather in financial performance itself.

Even if both selectivity and timing have true global means of zero, as theory requires, and also have finite-sample means of zero (which theory does not require), successful individual timers and/or selectors can still exist, provided the true variances of selectivity and timing are positive (i.e. some players win better than randomly off others, within the sample, in a zero sum game). The successes may or may not be longitudinally persistent over time. Our present results suggest that the true variances of ability between managers are probably small, and of low persistence between samples. However any true exceptions must be rare and small, if they are to have escaped universal recognition and elimination. A time-persistent variance need not be large to have a significant impact on the investors concerned. The true variances (and covariances) of selectivity and timing therefore remain keenly interesting, and as Coggin [5] remarked, a "fertile area" for theoretical and empirical study.

NOTES

1. Professor Charles Ward has told us verbally of high negative correlation when the Henriksson [9] model is used on Japanese market data.
2. This assumption is easily relaxed in the generating model, though it can pose difficulties for estimation. We model unavoidable empirical correlation between estimates of these components in equation (12).
3. Again this assumption is easily relaxed; we were surprised when our empirical data later appeared to violate this independence assumption, as explained below.
4. The foregoing structure departs from equation (1) by omitting variables X(t) and Y(t). In an existing empirical sample, an observed value of X(t) or Y(t) acts only as a scaling constant, which relates a regression coefficient to an element of monthly return. The scaling effect is allowed for later.
5. The notation in equations (15) to (17) below is not related to that for (1) above, so that no relationship is implied between X, Y and Z in (15) to (17) and X(t) and Y(t) in (1).
6. This requires non zero X, which always true in empirical work; similarly for Y and Z.
7. The use of $\underline{X}(t)^2$ and $\underline{Y}(t)^2$ as scaling constants assumes, as regression estimation itself does, that within any estimation sample the values of the beta coefficients are the random variables, and are conditional on constant, i.e. accurately observed, vector values of the independents X(t) and Y(t). The underlying statistical

generating model assumes the opposite, namely that the beta parameters are fixed, and the independent vectors variable. We call the observables X, Y, Z and their correlations the "input" parameters, and the estimated values of A, k, B_1, B_2, E_1 and E_2 the "output" parameters. This is the input-output sequence of calculation, which reverses the assumed causal sequence.

8. The term "degree of freedom" is used here in the sense of possible joint variation by parameters, not in the sense of the statistical sample size for a parameter. A parameter set with "one degree of freedom" in this sense is simply a scalar variable, perhaps in disguise, but there may be any number, from zero to infinity, of sampling degrees of freedom - i.e. of linearly independent sample points - available to estimate this scalar

REFERENCES

[1] Admati, A.R., S. Bhattacharya, P. Pfleiderer and S.A. Ross. (1986). "On Timing and Selectivity", with discussion by Verrechia, R.E., *Journal of Finance*, vol. XLI, 715-732.

[2] Armada, M.J. Rocha. (1990). "On the investigation of Selectivity and Market Timing in Portfolio Management", Paper presented at the Doctoral Tutorial, European Finance Association Conference, Athens, (August-September).

[3] Armada, M.J. Rocha. (1992). "On the investigation of Selectivity and Market Timing in Portfolio Management", PhD dissertation, Manchester Business School, University of Manchester.

[4] Chang, E. and W.G. Llewellen. (1984). "Market Timing and Mutual Fund Investment Performance", *Journal of Business*, vol. 57, 57-72.

[5] Coggin, T.D., F.J. Fabozzi and S. Rahman. (1993). "The investment performance of equity pension fund managers: an empirical investigation." *Journal of Finance*, vol. XXLVIII, 3, 1039-1055.

[6] Coggin, T.D. and J.E. Hunter. (1993). "A meta-analysis of mutual fund performance", *Review of Quantitative Finance and Accounting*, vol. 3, 189-201.

[7] Connor, G. and Korajczyk. (1991). "The attributes, behavior, and performance of US mutual funds", *Review of Quantitative Finance and Accounting*, 1, 5-26.

[8] Henriksson, R.D. (1984). "Market Timing and Mutual Fund performance: an empirical investigation", *Journal of Business*, vol. 57, 73-96.
[9] Henriksson, R.D. and R.C. Merton. (1981). "On Market Timing and investment performance II: Statistical Procedures for Evaluating Forecasting skills", *Journal of Business*, 54, 513-533.
[10] Hunter, J.E., T.D. Coggin and S. Rahman. (1992). " The correlation between sampling errors in estimating security selection and market timing ability" ,Working paper, Michigan State University, East Lansing, MI.
[11] Jagannathan, R. and R.A. Korajczyk. (1986). "Assessing the Market Timing performance of managed portfolios", *Journal of Business*, vol. 59, 217-236.
[12] Kon, Stayley J. (1983). "The Market Timing performance of Mutual Fund Managers", *Journal of Business,* vol. 56, 323-347.
[13] Kon, Stayley J. and F.C. Jen. (1979). "The investment performance of Mutual Funds: an empirical investigation of Timing, Selectivity and Market efficiency", *Journal of Business*, vol. 52, 263-289.
[14] Lee, C.F. and S. Rahman. (1990). "Market Timing, Selectivity and Mutual Fund performance: an empirical investigation", *Journal of Business*, vol. 63, 261-278.
[15] Lehman, B.N. and D.M. Modest. (1987). "Mutual Fund performance evaluation: a comparison of benchmarks and benchmark comparisons", *Journal of Finance*, 52, 233-265.
[16] Merton, R.C. (1981). "On Market Timing and investment performance: An equilibrium theory of value for Market Forecasts", *Journal of Business*, vol. 54, 363-406.
[17] Pfleiderer, P. and S. Bhattacharya. (1983). "A note on performance evaluation", Technical report 714, Stanford University.
[18] Press S.J. (1972). *Applied Multivariate Analysis*, Holt Rinehart Winston.
[19] Thompson, R. (1973). "The effects of the optimal utilisation of market forecasts on Mutual Fund performance in the period 1960-69", Working Paper, Rochester, NY, University of Rochester Graduate School of Business.
[20] Treynor, J. and F. Black. (1973). "How to use security analysis to improve portfolio selectivity", *Journal of Business*, vol. 46, 66-86.

[21] Treynor, J.L. and K. Mazuy. (1966). "Can mutual funds outguess the market?”, *Harvard Business Review*, 44, 131-136.
[22] Wonnacott, R.J. and T.H. Wonnacott. (1970). "*Econometrics*", Wiley International.

APPENDIX

The effect of adding an explicit error term to the model of (14)

Assume an error term e of variance E added to total investment performance, so that the variance of total return t is A + B_1+ B_2 + E. Assume that this error term disturbs the estimate of b_1 so that:

$$\left(\hat{b}_1\right) = b_1 + e_1 + e$$

and

$$\hat{b}_2 = kb_1 + k + ke + b_2 + e_2$$

and

$$\hat{a} = t - \hat{b}_1 - \hat{b}_2$$

The resulting variance-covariance matrix between the three estimates, in the order $\hat{b}_1$ and $\hat{b}_2$ and $\hat{a}$ is:

$$\begin{bmatrix} B_1 + E_1 + E & kB_1 + kE_1 + kE & -kB_1 - kE_1 - kE \\ kB_1 + kE_1 + kE & k^2B_1 + B_2 k^2 E_1 + E_2 + k^2E & -k^2B_1 - k^2E_1 - k^2E \\ -kB_1 - kE_1 - kE & -k^2B_1 - k^2E_1 - k^2E & A + k^2B_1 + (k+1)^2E_1 + E_2 + k^2E \end{bmatrix}$$

Comparison with (14) shows the covariations have the expected pattern of positive and negative signs, and all of the other coefficients (except for k which is artifactual) are reduced in size by the presence of E. However E cannot be identified from the set of six equations in seven unknowns that are obtained by equating the above matrix to the estimated variance-covariance matrix of the regression coefficients.