THE ANOMALY OF SIZE: DOES IT REALLY MATTER?†

by

Ian M. Dobbs*

* Department of Accounting and Finance,
University of Newcastle upon Tyne, NE1 7RU, UK.

† Thanks are due to the anonymous referees for comments which have materially improved this paper. The usual disclaimer applies.

ABSTRACT

Recent work on the ‘size effect’ suggests that size-related regularities in asset prices (such as size, leverage, book to market equity etc.) should not be regarded as anomalies. This paper first clarifies the argument (by showing why the OLS cross section regression incorporating size related variables is necessarily misspecified) and follows this by assessing the likely quantitative magnitude of this type of bias in a simulation study calibrated on US data.

1. INTRODUCTION

What determines expected stock returns? Simple unconditional capital asset pricing models (Sharpe [1964], Lintner [1965], Black [1972] etc.) suggest that it is market beta, whilst the empirical evidence suggests beta is at best a minor determinant, and that size or size related[1] variables have a significant role to play[2] (although there is still empirical dispute over which variables can be regarded as 'independent determinants' of expected stock returns). In a recent and influential paper, Jonathan Berk [1995] has argued that the size effect (and by implication, possibly all the above size-related regularities) in asset prices should not be regarded as anomalies. In particular, he argues

> " (1) that size-related regularities should be observed in the economy, and
>
> (2) [that] size will in general explain the part of the cross-section of expected returns left unexplained by an incorrectly specified asset pricing model."
>
> (Berk [1995, p. 275])

The object of the present paper is to first clarify Berk's proposition that "size related regularities should be observed in the economy"; in our view, the correct interpretation of his analysis is that "a misspecified econometric model may make it *appear* that there are size related regularities even when none exist." This reinterpretation, if correct, matters since it suggests that, if the object is to determine whether there is a 'real' size effect, then either one should

(i) re-specify the econometric model, or

(ii) explore the extent to which the misspecification matters (as with beta measurement error, if the impact is found to be small, it can be ignored).

Section 2 follows Berk's analysis in using a simple example to clarify why it is that incorporating size related variables on the right hand side of a cross section regression necessarily implies a violation of the assumption that the error term should be uncorrelated with the explanatory variables.[3] Section 3 then extends the model to a more 'real world' case where there is noise in the estimates of both the portfolio betas and expected returns. Finally, section 4 presents a simulation study, calibrated using the data set from a recent study by Jagannathan and Wang [1995]. The object of this study is to explore the extent to which size related misspecification is likely to show through in practice.

2. A SIMPLE BIVARIATE CASE

To establish the argument more formally, and for comparison purposes, the framework used in Berk [1995] is adopted. A brief outline is as follows. The 'economy' comprises a set of assets, *I*, which generate end-of-period random cash flow $\tilde{p}_{it}$, $i \in I$. The assets are traded at time zero in a spot market, at prices $p_{i0}, i \in I$. The continuously compounded return on the time interval $[0,t]$ is given by $\tilde{r}_i = \ln(\tilde{p}_{it} / p_{i0})$ and the expected return is denoted $R_i = E_0(\tilde{r}_i)$.[4] Each asset is parameterised by the expected value of (log) cash flow, which is denoted $C_i = E_0(\ln \tilde{p}_{it})$. Given there is a cross sectional distribution of assets in this economy, the values $R_i, C_i \;\; i \in I$ can be regarded as realisations of (cross-section) random variables *R, C.*[5] Berk's argument is presented in its simplest form in the context of a bivariate regression of expected return against (logarithm of) size:

$$R_i = \gamma_0 + \gamma_1 P_{i0} + \varepsilon_i, \qquad (1)$$

where $P_{i0} = \ln(p_{i0})$. If this regression is "run" in the whole population, he demonstrates that, if the one period CAPM holds, there is, logically, a size effect. The argument is based on the assumption that, in the population, $\text{cov}(C,R) < \text{var}(R)$, and the fact that this can be shown to imply that the value of the (population) regression coefficient on size is negative. He thus argues "that size-related regularities should be observed in the economy" (Berk [1995, p. 275]).

In what follows, a particular ***interpretation*** of this result is emphasised, namely that 'market value merely ***appears*** to explain expected return'. The point is that the classical assumptions which underpin OLS regression do not hold for the cross section regression in (1) since the error term is correlated with the explanatory variable. As a consequence, the OLS estimator of γ_1 in equation (1) is biased and inconsistent; such a regression may show a 'size effect' even if there is none in the underlying data generation process. To see this, take (1) as the true generating process and focus on the 'population' estimator

$$\hat{\gamma}_1 = \text{cov}(R, P_0) / \text{var}(P_0) . \qquad (2)$$

Using (1),

$$\hat{\alpha}_1 = \frac{\text{cov}(R, P_0)}{\text{var}(P_0)} = \frac{\text{cov}(\gamma_0 + \gamma_1 P_0 + \varepsilon, P_0)}{\text{var}(P_0)} = \gamma_1 + \frac{\text{cov}(\varepsilon, P_0)}{\text{var}(P_0)} . \qquad (3)$$

The usual assumption in OLS regression is that the error term is independent of the exogenous variables (that is, $\text{cov}(\varepsilon, P_0)$ =0). Here this is not the case because by definition

$$R = C - P_0 . \qquad (4)$$

The key point is that whilst *C* is an exogenous variable, P_0 is not; indeed it is P_0 that is the real dependent variable, as can be seen by substituting (4) into (1). This gives

$$P_0 = (1+\gamma_1)^{-1}(C - \gamma_0 - \varepsilon) . \qquad (5)$$

Assuming (as in Berk [1995]) that the error term ε is independent of the exogenous random variable, *C*, so that $\text{cov}(C, \varepsilon) = 0$, it then follows that

$$\text{cov}(\varepsilon, P_0) = \text{cov}\left(\varepsilon, (1+\gamma_1)^{-1}(C - \gamma_0 - \varepsilon)\right) = -(1+\gamma_1)^{-1}\, \text{var}(\varepsilon) . \qquad (6)$$

Hence $\hat{\gamma}_1$ is biased:

$$\hat{\gamma}_1 = \gamma_1 - (1+\gamma_1)^{-1} \frac{\text{var}(\varepsilon)}{\text{var}(P_0)} . \qquad (7)$$

Now suppose that the true return generating process is such that expected returns are actually independent of size. That is, in (1), $\gamma_1 = 0$. Then the OLS estimate of γ_1 is simply

$$\hat{\gamma}_1 = -\text{var}(\varepsilon) / \text{var}(P_0) < 0 \qquad (8)$$

The above analysis makes it clear that, although there is no real size effect, a spurious 'size effect' naturally arises out of bias in estimation. Furthermore, if there ***is*** a real (negative) size effect, the induced bias will tend to make the effect appear larger than it really is.[6]

3. SIZE AND BETA IN CROSS SECTION REGRESSION

Section 2 established that bias in estimation could give rise to the appearance of a spurious size (or size related) effect even though none existed in the underlying generating process. This section shows that this observation applies to any cross section regression which involves a measure for expected return regressed on explanatory variables some of which are size related (size , leverage, dividend yield, earnings per share etc.), ***if*** there are other determinants of expected return which are omitted from the model specification. It also shows that if, by contrast, expected returns are generated by a single factor (beta) process, then there is ***no*** bias, so long as the cross section regression includes beta as a right hand side explanatory variable.

Suppose that, on the interval [0,*t*], the true relationship between expected returns, size and beta is as follows:

$$R_i = \alpha_0 + \alpha_1 \beta_i^* + \alpha_2 P_{i0} + \varepsilon_i \tag{9}$$

where β_i^* denotes the i^{th} asset's 'true' equity beta and as before, $P_{i0} = \ln(p_{i0})$ is the associated size variable. It is assumed that there is no size effect (so $\alpha_2 = 0$; size is included in the equation only because the object is to consider what happens if a regression of this type is run with the object of obtaining an estimate for α_2). Now, if the CAPM holds, not only must $\alpha_2 = 0$ but it must also be the case that $\varepsilon_i = 0$ for all $i \in I$. The error term ε_i can only deviate from zero if there are factors (other than beta) which influence expected return. In what follows, it is shown *inter alia* that, with beta as the sole factor generating expected returns, there is no 'spurious size effect' arising out of adding size as a right hand variable.

However, it can be argued that the point behind cross section testing of the CAPM is not so much to test whether the CAPM is 'true' (after all, there are many other predictions of the CAPM which are clearly and palpably false) as to simply test whether beta is a significant explanator of asset returns (and whether it does more or less work than other potential explanatory variables). The point of (our reinterpretation of) Berk's analysis is that, admitting there may be unspecified (non-size) determinants of *expected return* not included in the estimation model, the inclusion of a size variable on the right hand side causes a breakdown in the OLS assumption that the error is uncorrelated with this.

To move toward a more realistic model, the fact that beta is typically only estimated with error, and that actual rather than expected return is observed, are now incorporated. That the coefficients in a regression of type (9) will generally be biased because of 'errors in variables' (the fact that the equity betas are only estimated, and hence are best interpreted as being measured with error) has been known for some time (see e.g. Levi [1973], Fama and MacBeth [1973]) along with the fact that this will generate a spurious 'size effect' (see e.g. Shanken [1985], Chan and Chen [1988], Kothari, Shanken and Sloan [1995]) although the effect is, empirically, fairly small when, as is customary, the cross section analysis is conducted using portfolios rather than individual assets. The analysis in section 2 indicates that the size variable can create an additional and potentially much more substantial source of bias.

The rest of this section is devoted to obtaining expressions for the extent of bias in the 'population regression' OLS estimators of α_1, α_2. The observed beta is given as

$$\beta_i = \beta_i^* + \psi_i \qquad (10)$$

where ψ_i denotes random measurement error. Finally, the observed return can be written as

$$r_i = R_i + \nu_i \qquad (11)$$

where ν_i simply reflects the fact that the end of period price may deviate from its time zero expectation.

Thus (9) can be rewritten as

$$r_i = \alpha_0 + \alpha_1 \beta_i + \alpha_2 P_{i0} + \eta_i \tag{12}$$

where

$$\eta_i = \varepsilon_i - \alpha_1 \psi_i + \nu_i \,. \tag{13}$$

The OLS estimator for α_2 in (12) is thus

$$\hat{\alpha}_2 = \frac{\text{cov}(P_0, r)\,\text{var}(\beta) - \text{cov}(\beta, r)\,\text{cov}(\beta, P_0)}{\text{var}(\beta)\,\text{var}(P_0) - \text{cov}(\beta, P_0)^2}\,. \tag{14}$$

(where the covariances etc. are evaluated across the assets $i \in I$). As already discussed in section 2, the real dependent variable in the system (10)-(13) is P_0, whilst the variables $\beta^*, C\left(\equiv E_0(P_t)\right)$ are exogenous. In particular, from (4) and (11),

$$r = R + \nu = C - P_0 + \nu \tag{15}$$

The appropriate assumptions regarding the error terms include

$$\text{cov}(\beta^*, \varepsilon), \text{cov}(\beta^*, \psi), \text{cov}(C, \varepsilon), \text{cov}(C, \psi), \text{cov}(\psi, \varepsilon) = 0\,.$$

$$\text{cov}(\beta^*, \nu), \text{cov}(C, \nu), \text{cov}(\nu, \psi), \text{cov}(\nu, \varepsilon) = 0$$

Given this, it is straightforward to show that (see appendix)

$$\hat{\alpha}_2 - \alpha_2 = \frac{\alpha_1\,\text{var}(\psi)\,\text{cov}(\beta, P_0) - (1+\alpha_2)^{-1}\,\text{var}(\xi)\,\text{var}(\beta)}{\text{var}(\beta)\,\text{var}(P_0) - \text{cov}(\beta, P_0)^2}\,. \tag{16}$$

where

$$\text{var}(\xi) = \text{var}(\varepsilon - \alpha_1 \psi) = \text{var}(\varepsilon) + \alpha_1^2\,\text{var}(\psi) \tag{17}$$

Thus, under the null hypothesis that there is no real size effect, such that $\alpha_2 = 0$, then

$$\hat{\alpha}_2 = \frac{\overbrace{\alpha_1}^{(+)}\overbrace{\text{var}(\psi)}^{(+)}\overbrace{\text{cov}(\beta, P_0)}^{(-)} - \overbrace{\text{var}(\xi)}^{(+)}\overbrace{\text{var}(\beta)}^{(+)}}{\underbrace{\text{var}(\beta)\,\text{var}(P_0) - \text{cov}(\beta, P_0)^2}_{(+)}}\,. \tag{18}$$

The equivalent calculation for $\hat{\alpha}_1$ (again setting $\alpha_2 = 0$) gives

$$\hat{\alpha}_1 - \alpha_1 = -\frac{\overbrace{\alpha_1}^{(+)}\overbrace{\text{var}(\psi)}^{(+)}\overbrace{\text{var}(P_0)}^{(+)} - \overbrace{\text{var}(\xi)}^{(+)}\overbrace{\text{cov}(\beta, P_0)}^{(-)}}{\underbrace{\text{var}(\beta)\,\text{var}(P_0) - \text{cov}(\beta, P_0)^2}_{(+)}}\,. \tag{19}$$

Clearly, $\text{var}(\psi), \text{var}(\varepsilon), \text{var}(\xi) > 0$, and it is usually argued that, empirically, size and beta are negatively correlated, so $\text{cov}(\beta, P_0) < 0$. Finally, the assumption that investors are risk averse implies $\alpha_1 > 0$. The denominator of the ratio in (18) is positive, so, with the numerator negative, this implies the 'illusion' of a size effect even when there isn't one:

$$\alpha_2 = 0 \;\Rightarrow\; \hat{\alpha}_2 < 0. \tag{20}$$

The effect of 'errors in variables' was already known to give a spurious negative 'size effect' and the extent of this bias is increased further by the 'correlated error' effect. Turning now to $\hat{\alpha}_1$, both effects again operate in the same direction, giving

$$\hat{\alpha}_1 < \alpha_1. \tag{21}$$

It follows that empirical analysis which includes a size variable increases the likelihood that beta will ***appear*** not to be a significant factor.

The 'spurious size effect' is thus manifest if either

(i) betas are measured with error, and/or

(ii) there are other (non-size) related determinants of expected returns.

However, clearly, if there is no beta measurement error ***and*** the CAPM holds (only beta determines expected return), then there is no bias. In this case, $\text{var}(\psi) = 0$, $\text{var}(\varepsilon) = 0$ and hence $\text{var}(\xi) = \text{var}(\varepsilon - \alpha_1\psi) = 0$. Hence $\hat{\alpha}_1 = \alpha_1$ and $\hat{\alpha}_2 = \alpha_2 = 0$.

The above analysis suggests that, with a multi-factor generating process, the observation of size effects could be merely an artefact of a misspecified statistical model. The inclusion of size variables on the right hand side in such circumstances also increases the probability that beta will not appear as a significant factor. A natural question to pose then is whether such effects are likely to be significant in empirical applications. This question is addressed in the simulation study in section 4 below.

4. ASSESSING THE EXTENT OF BIAS: A SIMULATION STUDY

This section explores the potential quantitative impact of the above form of specification error. The simulation is 'calibrated' using a US data set comprising 100 portfolios constructed from NYSE and NASDAQ stocks as used in recent work on capital asset pricing models by Jagannathan and Wang [1995] (JW hereafter). This data set gives an initial ball park for portfolio characteristics (beta, size, unsystematic risk etc.) typical for cross section studies of this type. Where appropriate, the study uses sensitivity analysis to consider variations from these characteristics (for example, in portfolio total risk, or for measurement error in beta etc.).

The JW data set comprises returns on 100 NYSE and NASDAQ stock portfolios constructed using the Fama and French [1992] sorting procedure (for construction details, see JW [1995, pages 18-19][7]). The sorting procedure generates portfolios which differ markedly according to size and beta. For each portfolio, the monthly returns from July 1963-December 1990 give a time series comprising 330 observations. The beta coefficients reported below were computed by regressing each portfolio return against the return (denoted R_{mt}) on the CRSP value weighted index for the corresponding months. That is,

$$R_{it} = \alpha_i + \beta_i R_{mt} + \eta_{it} \qquad (t=1,..,330) \qquad (i=1,..,100) \qquad (22)$$

The average monthly returns on these portfolios are reported in table 1. Table 2 reports the estimates $\hat{\beta}_i$, i=1,..,100 for the portfolios, whilst table 3 gives the portfolio size variable. Also recovered from this set of regressions are estimates for the standard error for beta, $s_{\hat{\beta}_i}$ (table 4) and unsystematic risk, s_{η_i} (table 5).[8]

Tables 1-6 about here

A typical cross section regression using this type of data set involves calculating the average return for each portfolio and regressing this against the portfolio beta and other cross sectional explanatory variables. Table 6 reports the results obtained[9] for the simple regressions of (a) return on beta and (b) return against beta and size. The evidence here is that, in regression (a), the coefficient on beta is not significantly different from zero, whilst in (b), introducing size, both beta and size are significantly negative. This result is consistent with the effect of correlated error bias as examined in section 3.

The simulation model is specified as follows. The estimated betas, standard errors, unsystematic risks etc. from the times series regressions in (22) are used to 'calibrate' the data generation process. That is, set $\beta_i{*} = \hat{\beta}_i$, $\text{var}(\psi_i) = s^2_{\hat{\beta}_i}$, $\text{var}(\eta_i) = s^2_{\eta_i}$ and use the size variable reported in JW (table 3 here) to characterise the exogenous variable $E_0(P_t)$. Clearly the time series estimate $s^2_{\hat{\beta}_i}$ is likely to be an over-estimate for $\text{var}(\psi_i)$, whilst in different studies using different portfolios may feature different general levels of unsystematic risk. This is taken into account by conducting a sensitivity analysis in which the impact of variations in both $\text{var}(\psi_i)$ and $\text{var}(\eta_i)$ are examined (a general conclusion, as one might expect, is that even quite large variations in $\text{var}(\psi_i)$ have little impact on the level of bias in the beta and size parameters).

The simulation involves generating $n = 1,..n_{reps}$ (=5,000) simulated data sets. Each data set is generated as follows. Observed betas are generated according to the process

$$\beta_i(n) = \beta_i{*} + \psi_i(n) \qquad \psi_i(n) \sim N(0, \text{var}(\psi_i)) \qquad (i=1,..,100) \qquad (23)$$

Thus the estimated betas are used as values for $\beta_i *$ and their standard errors used to generate 'observed' betas, denoted $\beta_i(n)$. Expected portfolio returns are then generated using

$$R_i(n) = \alpha_0 + \alpha_1\beta_i * + \varepsilon_i(n) \qquad \varepsilon_i(n) \sim N(0, \mathrm{var}(\varepsilon_i)) \quad (i{=}1,..,100) \qquad (24)$$

after which observed returns are generated by

$$r_i(n) = R_i(n) + \nu_i(n) \qquad \nu_i(n) \sim N(0, \mathrm{var}(\nu_i)) \qquad (i{=}1,..,100) \qquad (25)$$

Unfortunately, on the basis of the empirical data, it is not possible to completely disentangle the variances for ε, ψ, ν. What we do know is that, since $\eta_i = \varepsilon_i - \alpha_1\psi_i + \nu_i$ and since ε, ψ, ν are independent random variables, it follows that

$$\mathrm{var}(\varepsilon_i) + \mathrm{var}(\nu_i) = \mathrm{var}(\eta_i) - \alpha_1^2\,\mathrm{var}(\psi_i)\,, \qquad (i{=}1,..,100) \qquad (26)$$

and we do have estimates for the right hand side variances. A useful way to proceed is through sensitivity analysis. Accordingly, the model is parameterised as follows. We set

$$\mathrm{var}(\eta_i) = k_1 s_{\eta_i}^2 \qquad (i{=}1,..,100) \qquad (27)$$

$$\mathrm{var}(\psi_i) = k_2 s_{\hat{\beta}_i}^2 \qquad (i{=}1,..,100) \qquad (28)$$

$$\mathrm{var}(\varepsilon_i) = k_3[\mathrm{var}(\eta_i) - \alpha_1^2\,\mathrm{var}(\psi_i)] \qquad (i{=}1,..,100) \qquad (29)$$

$$\mathrm{var}(\nu_i) = (1 - k_3)[\mathrm{var}(\eta_i) - \alpha_1^2\,\mathrm{var}(\psi_i)] \qquad (i{=}1,..,100) \qquad (30)$$

The sensitivity analysis then involves an assessment of the impact of altering each of $k_1,..,k_4$. Thus, varying k_1 allows an exploration of the impact of varying total noise (k_1=1 gives portfolio total variances as estimated in JW data); varying k_2 allows an exploration of the impact of varying the overall level of noise in the estimate of beta (k_2 =0 implies betas estimated without error, whilst k_2 =1 gives each portfolio beta a standard error equal to that in the JW data). The parameter k_3 acts as a 'variance splitter' ($0 \le k_3 \le 1$) which divides the total variance on the right hand side in (26) between that associated with determinants of expected return ($\mathrm{var}(\varepsilon)$) and that associated with the deviation of actual from expected return ($\mathrm{var}(\nu)$) (thus setting k_3=1 gives maximum bias).

The values of α_0, α_1, respectively the risk free rate and the market premium in the Sharpe CAPM, can also be expected to have an effect on bias, and so are also treated as parameters in the simulation. Denoting the observed total return on the market index used in the JW data set over this period as R_m, then the values of α_0, α_1 are selected to satisfy

$$R_m = \alpha_0 + \alpha_1 \beta_m = \alpha_0 + \alpha_1, \qquad (31)$$

so that, given a choice of α_1, the value of α_0 is determined as

$$\alpha_0 = R_m - \alpha_1 \qquad (32)$$

We therefore parameterise α_0, α_1 by setting

$$\alpha_1 = k_4 R_m \qquad (33)$$

with $0 \le k_4 \le 1$, with α_0 then defined by (32). Thus setting $k_4 = 0$ implies a data generating process independent of beta etc.

For given values of $k_1,..,k_4$, equations (23), (24), (25) are used to generate a cross section data set for $r_i(n), \beta_i(n), P_{i0}(n)$ (i=1,..,100). The endogenous variable $P_{i0}(n)$ is generated using (24) and the fact that $P_{i0}(n) = E_0(P_{it}) - R_i(n)$ (from (4)). Following this, the cross section regression

$$r_i(n) = a_0 + a_1 \beta_i(n) + a_2 P_{i0}(n) + e_i(n) \qquad (i=1,..,100) \qquad (34)$$

is estimated. The whole process is repeated (for fixed $k_1,..,k_4$) $n = 1,..,n_{reps}$ (=5,000)[10] times in order to generate the finite sample characteristics of the OLS estimators for α_0, α_1. The 'population' bias values are also calculated, using equations (18) and (19). Finally, the values of $k_1,..,k_4$ are varied to explore their impact on finite sample and population bias measures. The results are reported in table 7 (for the size parameter) and 8 (for beta).

Table 7, 8 here

In table 7, column 6 gives population bias, whilst columns 7-11 report descriptive statistics for the size parameter; thus column 7 reports the average level of bias over the 5000 finite sample regressions, whilst column 8 gives the associated average *t*-value for the size parameter. Columns 9-11 give the percentage of the total number of regressions in which the size parameter is significant at the 10%, 5%, 2.5% levels of significance (one tailed test). Table 8, focussing on the beta parameter a_1 gives similar statistics, although we also report in column 6 the 'true' value of α_1 used in the simulation (since this varies across cases). The average value of the *t*-statistic is reported both for the null hypothesis that 0 is the true value (in column 9) and also against the true value of α_1 as used in the data generation process (column 10). Again, columns 11-13 report the number of regressions in which $\hat{\alpha}_1$ is significantly greater than zero (one tailed test) at the 10, 5 and 2.5% levels of significance.

The cases are organised as follows. Setting $k_1 = 0.5$ gives a value for α_1 equal to half the market rate R_m (in a Sharpe model, equivalent to setting the market premium equal to the risk free rate). The first three cases (both tables) then set simulation portfolio variances equal to those found in the JW data (i.e. $k_2 = 1$), with half the noise being ascribed to expected return variance and half to actual return (i.e. $k_3 = 0.5 \Rightarrow \text{var}(\varepsilon_i) = \text{var}(\nu_i)$). Finally $\text{var}(\psi_i)$ is increased from zero (case 1) through to the values estimated in the JW data (case 2) and then to 10 times these (case 3). The results indicate (the well known result) that, at the level of beta estimation noise typical in portfolio analysis, bias due to errors in beta has little impact on bias.

Having explored the impact of beta-estimation bias, and given that it is relatively small in any case, this is put to one side in cases 4-11. Cases 4 and 5 indicate just how important to bias is the value of k_3; recall that when k_3=0, this sets $\text{var}(\varepsilon_i) = 0$, whilst k_3=1 sets $\text{var}(\nu_i) = 0$, such that the bias effect is at its maximum (for given total variance). Thus case 4 merely illustrates the fact that there is no bias at all in population or finite sample estimators when $\text{var}(\varepsilon_i) = \text{var}(\psi_i) = 0$ (for all *i*=1,..,100), whilst there is a highly significant impact when $\text{var}(\varepsilon_i)$ is set at its maximum value (the average *t*-value on size in case 5 being -4.27).

Cases 6,7 illustrate the impact on bias of varying the total volatility of the portfolios. Setting k_2 =0.5 cuts the total variance for each portfolio to one half of that observed in the JW data, whilst k_2 =5 multiplies it five fold. The impact is as expected; notice that, in table 8, the strength of the effect in case 7 is sufficient to make beta very rarely significantly positive.

Cases 9-11 explore the impact of varying the true beta coefficient, from $\alpha_1 = 0$ through to $\alpha_1 = R_m$. This has little impact on the bias observed in the size coefficient $\hat{\alpha}_2$, but does of course have an impact on the ability of the regression to pick up a significant beta effect. When the beta effect is relatively small (10 or 20% of R_m in cases 9, 10 in table 8), the bias impact is sufficient to leave the coefficient insignificant in most regressions.

Finally, in view of the fact that in many cross section regressions recently reported, beta is observed to have no significant impact, cases 12 and 13 set $\alpha_1 = 0$ and total variance equal to that observed in the JW data, with standard errors on betas as in that data; the tables here give the impact of varying expected return noise from its maximum (case 12) to its minimum of 0 (case 13). As expected, the size related bias effect shows through heavily in case 12. However, it is worth noting that, when true α_1=0, the bias on this coefficient is relatively smaller than when α_1 is larger (as expected, given equations (16) and (18)).

To sum up, the simulation reveals that the extent of bias in finite sample cross section regression depends upon the proportion of variance in (26) assigned to $\text{var}(\varepsilon_i)$. At its maximum (when $k_3 = 1$), the bias can be highly significant in cross section regression, to the extent that it could easily result in incorrect inferences on both the contribution of size and beta. In the present framework, it is not possible to disentangle the two sources of variation in actual return ($\text{var}(\varepsilon_i)$ and $\text{var}(\nu_i)$), so it is not possible to say which is the more important.[11] However, the above analysis does point to the fact that in practical applications the observation of a significant size effect *could* well be the result of correlated error bias.

5. CONCLUDING COMMENTS

The object of this paper has been to clarify the debate as to whether the observation of size effects in cross section regression analysis constitutes an ‘anomaly’ or not. We have argued that with beta as the only factor in the data generating process, there is no problem with adding a size variable to the right hand side of the cross section regression; under the null hypothesis that the CAPM is ‘true’, the estimators of the size and beta effects are unbiased. However, if the underlying data generating process is multi-factor, including non-size factors, then any cross section regression which includes beta and size as explanatory variables (but excludes other explanatory variables), necessarily leads to correlated error bias in the parameter estimators for both size and beta. The effect is to increase the chance that, in a finite sample regression, size will appear to be significant when it is not - and that beta will be more likely to appear with an insignificant coefficient even if does have real explanatory power. The evidence provided by the simulation study suggests that in typical cross section studies, the bias-effect *could* be important. Although the focus has been exclusively on size, similar effects occur in cross section regressions where (market value based) leverage, dividend yield, book to market equity, earnings yield, etc. are included on the right hand side.

R E F E R E N C E S

Banz R.W., 1981, The relationship between return and market value of common stocks, Journal of Financial Economics, 9, 3-18.

Basu S., 1983, The relationship between earnings yield, market value and return for NYSE common stocks: Further evidence, Journal of Financial Economics, 12, 129-156.

Bhandari L.C., 1988, Debt/equity ratio and expected common stock returns: Empirical evidence, Journal of Finance, 43, 507-528.

Berk J., 1995, A critique of size related anomalies, Review of Financial Studies, 8, 275-286.

Black F., 1972, Capital market equilibrium with restricted borrowing, Journal of Business, 45, 444-455.

Chan K. and N. Chen, 1988, An unconditional asset pricing test and the role of firm size as an instrumental variable for risk, Journal of Finance, 43, 309-325.

Fama E. and K. French, 1992, The cross section of expected stock returns, Journal of Finance, 47, 427-465.

Fama E. and J. MacBeth, 1973, Risk, return and equilibrium: Empirical tests, Journal of Political Economy, 81, 607-636.

Handa P., Kothari S.P. and C. Wasley, 1989, The relation between the return interval and betas: Implications for the size effect, Journal of Financial Economics, 23, 79-100.

Jagannathan R. and Wang Z., 1996, The conditional CAPM and the cross-section of expected returns, Journal of Finance, 51, 3-53.

Kothari S.P., Shanken J. and R.G. Sloan, 1995 Another look at the cross-section of expected stock returns, Journal of Finance, 50, 185-224.

Levi M., 1973, Errors in variables bias in the presence of correctly measured variables, Econometrica, 41, 985-986.

Lintner J., 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, Review of Economics and Statistics, 47, 13-37.

MacKinlay A.C., 1995, Multifactor models do not explain deviations from the CAPM, Journal of Financial Economics, 38, 3-28.

Rosenberg B., Reid K., and R. Lanstein, 1985, Persuasive evidence of market inefficiency, Journal of Portfolio Management, 11, 9-17.

Shanken J., 1985, Multi-variate tests of the zero beta CAPM, Journal of Financial Economics, 14, 327-348.

Sharpe W.F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, Journal of Finance, 19, 425-442.

Table 1: Portfolio Average monthly returns

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
S1	1.44%	1.53%	1.56%	1.71%	1.36%	1.44%	1.37%	1.33%	1.46%	1.34%
S2	1.13%	1.22%	1.09%	1.19%	1.38%	1.37%	1.37%	1.30%	1.15%	0.95%
S3	1.26%	1.27%	1.22%	1.26%	1.16%	1.29%	1.34%	1.19%	1.12%	0.89%
S4	1.37%	1.47%	1.40%	1.28%	1.01%	1.39%	1.11%	1.33%	1.07%	0.95%
S5	0.97%	1.53%	1.10%	1.28%	1.18%	1.04%	1.35%	1.07%	1.23%	0.82%
S6	1.07%	1.36%	1.34%	1.12%	1.25%	1.27%	0.84%	0.94%	0.92%	0.77%
S7	0.99%	1.18%	1.13%	1.19%	0.96%	0.99%	1.11%	0.91%	0.90%	0.83%
S8	0.95%	1.19%	1.02%	1.39%	1.18%	1.24%	0.94%	1.02%	0.88%	1.08%
S9	0.94%	0.92%	1.05%	1.17%	1.15%	1.03%	1.02%	0.84%	0.80%	0.51%
S10	1.06%	0.97%	1.02%	0.94%	0.83%	0.93%	0.82%	0.83%	0.61%	0.72%

Table 2: Portfolio Betas

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
S1	0.90	0.99	1.01	1.13	1.17	1.21	1.2	1.31	1.44	1.54
S2	0.83	1.00	1.09	1.12	1.18	1.29	1.33	1.39	1.48	1.63
S3	0.78	0.93	1.09	1.11	1.18	1.27	1.29	1.40	1.42	1.70
S4	0.75	0.91	1.05	1.13	1.19	1.32	1.25	1.32	1.56	1.61
S5	0.57	0.78	1.10	1.10	1.12	1.20	1.25	1.43	1.45	1.54
S6	0.62	0.77	0.88	1.01	1.08	1.25	1.22	1.34	1.32	1.59
S7	0.64	0.84	1.01	1.07	1.16	1.21	1.26	1.26	1.31	1.54
S8	0.64	0.73	0.91	1.04	1.07	1.17	1.22	1.19	1.23	1.50
S9	0.62	0.78	0.88	0.96	1.04	1.05	1.13	1.17	1.22	1.34
S10	0.68	0.76	0.80	1.00	0.97	1.00	1.04	1.09	1.10	1.28

Table 3: Portfolio Size (log$M)

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
S1	2.48	2.5	2.49	2.48	2.48	2.50	2.46	2.46	2.46	2.34
S2	3.71	3.72	3.73	3.73	3.71	3.71	3.72	3.72	3.72	3.72
S3	4.21	4.21	4.21	4.21	4.21	4.23	4.21	4.22	4.21	4.20
S4	4.67	4.65	4.64	4.65	4.65	4.65	4.65	4.64	4.64	4.64
S5	5.07	5.09	5.07	5.08	5.08	5.07	5.07	5.07	5.07	5.05
S6	5.47	5.48	5.47	5.48	5.48	5.48	5.48	5.48	5.47	5.48
S7	5.91	5.92	5.93	5.92	5.92	5.89	5.91	5.90	5.92	5.90
S8	6.44	6.42	6.43	6.39	6.43	6.41	6.43	6.42	6.40	6.40
S9	6.98	6.98	7.00	6.98	6.96	6.97	6.95	6.96	6.95	6.97
S10	8.11	8.26	8.22	8.19	8.16	8.18	8.06	8.03	7.92	7.81

Table 4: $s_{\hat{\beta}}$, the standard error on portfolio beta

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
S1	0.054	0.054	0.058	0.062	0.064	0.064	0.067	0.070	0.075	0.082
S2	0.042	0.049	0.046	0.047	0.055	0.055	0.053	0.055	0.060	0.068
S3	0.041	0.042	0.047	0.046	0.050	0.048	0.052	0.051	0.052	0.060
S4	0.036	0.037	0.040	0.041	0.042	0.048	0.047	0.047	0.060	0.059
S5	0.036	0.034	0.031	0.036	0.040	0.044	0.041	0.045	0.050	0.054
S6	0.036	0.030	0.030	0.034	0.034	0.039	0.041	0.039	0.051	0.054
S7	0.037	0.035	0.030	0.030	0.034	0.032	0.034	0.035	0.039	0.049
S8	0.039	0.032	0.028	0.029	0.032	0.029	0.031	0.035	0.039	0.047
S9	0.037	0.029	0.026	0.028	0.029	0.028	0.031	0.030	0.030	0.038
S10	0.034	0.028	0.028	0.025	0.023	0.026	0.024	0.024	0.027	0.033

Table 5: $s_{\hat{\eta}}$, Portfolio unsystematic risk

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
S1	0.044	0.044	0.048	0.050	0.052	0.052	0.055	0.057	0.061	0.067
S2	0.034	0.040	0.037	0.038	0.045	0.045	0.043	0.045	0.049	0.055
S3	0.033	0.034	0.038	0.037	0.041	0.039	0.043	0.041	0.042	0.049
S4	0.029	0.030	0.032	0.033	0.035	0.039	0.039	0.038	0.049	0.048
S5	0.030	0.027	0.026	0.029	0.032	0.036	0.033	0.037	0.041	0.044
S6	0.029	0.025	0.025	0.028	0.028	0.032	0.033	0.032	0.042	0.044
S7	0.030	0.028	0.025	0.025	0.027	0.026	0.028	0.028	0.031	0.040
S8	0.032	0.026	0.023	0.024	0.026	0.023	0.025	0.028	0.032	0.039
S9	0.030	0.023	0.021	0.022	0.024	0.023	0.025	0.024	0.024	0.031
S10	0.027	0.023	0.023	0.020	0.019	0.021	0.020	0.020	0.022	0.027

Table 6: Cross Section Results for the original JW data

	Coefficients			
OLS Regression	Constant	Beta	Size	R-square
(a) on Beta	4.103 (12.15)	-0.338 (-1.16)	- (-)	0.0135
(b) on Beta and Size	6.858 (20.81)	-1.050 (-5.19)	-0.368 (-11.34)	0.5757

Table 7: Simulation Results, Size Parameter.

1	2	3	4	5	6	7	8	9	10	11
Case	k_1	k_2	k_3	k_4	Population bias in $\hat{\alpha}_2$	Average bias in finite sample estimator $\hat{\alpha}_2$	Average *t*-value (for null hypothesis that α_2 =0)	% of regressions in which the beta parameter $\hat{\alpha}_2$ is significantly less than zero at the significance level:		
								10%	5%	2.5%
1	0.5	1	0.5	0	-0.086	-0.084	-2.099	76.320	65.580	54.560
2	0.5	1	0.5	1	-0.090	-0.087	-2.175	78.780	67.860	57.080
3	0.5	1	0.5	10	-0.119	-0.144	-3.854	99.180	98.000	96.000
4	0.5	1	0	0	0.000	0.000	-0.008	11.680	6.260	3.320
5	0.5	1	1	0	-0.159	-0.154	-4.272	99.620	99.100	98.260
6	0.5	0.5	1	0	-0.086	-0.084	-3.012	94.420	89.300	82.960
7	0.5	5	1	0	-0.486	-0.479	-9.537	100.000	100.000	100.000
8	0	1	1	0	-0.159	-0.154	-4.238	99.640	99.220	98.440
9	0.1	1	1	0	-0.159	-0.154	-4.233	99.720	99.200	98.000
10	0.2	1	1	0	-0.159	-0.155	-4.280	99.800	99.260	98.240
11	1	1	1	0	-0.159	-0.155	-4.266	99.660	99.120	98.160
12	0	1	1	1	-0.158	-0.154	-4.261	99.680	99.300	97.980
13	0	1	0	1	0.000	0.001	0.024	11.140	6.320	3.200

k_1= value of α_1 as proportion of $\overline{R}_m$.

k_2 = total variance multiplier (k_2 =1 gives portfolio variances as in original JW data).

k_3 splits variance between ε and ν (k_3=1 gives maximum bias effect).

k_4 =multiplier on beta error variance (k_4 =1 implies betas have same standard errors as in original JW data; k_4 =0.5 halves this etc.).

Table 8: Simulation Results, Beta parameter

1	2	3	4	5	6	7	8	9	10	11	12	13
Case	k_1	k_2	k_3	k_4	α_1 (True value)	Population bias in $\hat{\alpha}_1$	Average bias in finite sample estimator $\hat{\alpha}_1$	Average *t*-value (for null hypothesis that α_1=0)	Average *t*-value (for null hypothesis that α_1 equals its true value)	% of regressions in which the beta parameter α_1 is significantly greater than zero at the significance level		
										10%	5%	2.5%
1	0.5	1	0.5	0	1.497	-0.296	-0.286	4.247	-1.009	99.820	99.440	98.680
2	0.5	1	0.5	1	1.497	-0.345	-0.336	4.147	-1.204	99.720	99.140	98.060
3	0.5	1	0.5	10	1.497	-0.659	-1.242	1.923	-9.354	71.060	59.940	48.520
4	0.5	1	0	0	1.497	0.000	0.004	5.077	0.011	99.960	99.940	99.780
5	0.5	1	1	0	1.497	-0.545	-0.534	3.612	-2.019	99.580	98.300	96.000
6	0.5	0.5	1	0	1.497	-0.296	-0.288	6.120	-1.466	100.000	100.000	100.000
7	0.5	5	1	0	1.497	-1.667	-1.643	-0.331	-3.669	2.280	0.720	0.260
8	0	1	1	0	0.000	-0.307	-0.304	-1.240	-1.240	0.400	0.120	0.080
9	0.1	1	1	0	0.299	-0.355	-0.347	-0.195	-1.397	6.980	2.720	1.180
10	0.2	1	1	0	0.599	-0.402	-0.394	0.802	-1.570	30.340	19.860	11.980
11	1	1	1	0	2.994	-0.783	-0.764	7.513	-2.588	100.000	100.000	100.000
12	0	1	1	1	0.000	-0.298	-0.290	-1.206	-1.206	0.780	0.240	0.080
13	0	1	0	1	0.000	0.000	0.002	0.009	0.009	12.080	6.580	3.460

FOOTNOTES

[1] These are termed ‘size related’ because they incorporate size either directly or in some form of ratio with other variables.

[2] Some selected references on the size related determinants of expected stock returns (the literature is massive):

(i) size (e.g. Banz [1981], Chan and Chen [1988], Handa, Kothari and Wasley [1989], Fama and French [1992])
(ii) leverage (e.g. Bhandari [1988])
(iii) book to market equity (e.g. Rosenberg, Reid and Lanstein [1985], Fama and French [1992, 1993])
(iv) earnings yield (e.g. Basu [1983]).

Much recent work on the determinants of stock returns has focussed on measurement problems (data snooping and survivorship bias etc.), market frictions, investor ‘irrationality’ (see e.g. MacKinlay [1995] for discussion) or variants of the conditional CAPM (e.g. Jagannathan and Wang [1995]).

[3] Section 2 essentially reworks his analysis to emphasise why it can be understood as a problem of misspecification and bias in estimation.

[4] Where $E_0(.)$ denotes that expectations are formed at time 0.

[5] Clearly the structure of the bi-variate distribution function for *C,R* is such that $\Pr(C = C_i) = \Pr(C = C_j)$ and $\Pr(R = R_i) = \Pr(R = R_j)$ for all $i, j \in I$ etc.

[6] For $\gamma_1 > -1$; absolute bias increases the more negative γ_1 is in this range.

[7] This data set is in fact available over the internet; see the paper by Jagannathan and Wang [1995] for details regarding how to download this data.

[8] Tables 1-3 give data/results also reported in Jagannathan and Wang [1995]; we replicate their work in order to add results for standard errors on betas and also for unsystematic risk (tables 4,5).

[9] Table 6 replicates results reported by JW (although the unconditional CAPM was not the focus of their study). The only difference between the results in table 6 and those of JW is that here the regression is of total return against beta and size, whilst JW report the result for average return against these variables. Given there are 330 observations, the difference amounts to a scale factor of 330 for parameter estimates; *R*-square, *t*-values etc. are naturally unaffected.

[10] This number of replications gives adequate stability for our purposes.

[11] Notice that using end of period size does not side-step the problem that the size variable is correlated with the error term; in this case the problem is with ν rather than ε.