THE JOURNAL OF FINANCE • VOL. LV, NO. 1 • FEBRUARY 2000

Sorting Out Sorts

JONATHAN B. BERK*

ABSTRACT

In this paper we analyze the theoretical implications of sorting data into groups and then running asset pricing tests within each group. We show that the way this procedure is implemented introduces a bias in favor of rejecting the model under consideration. By simply picking enough groups to sort into, the true asset pricing model can be shown to have no explanatory power within each group.

THE COMMON PRACTICE OF SORTING STOCKS into groups to test asset pricing inferences began with the earliest tests of the CAPM (see Black, Jensen, and Scholes (1972)). Although the information loss from the sorting procedure has long been recognized (see Litzenberger and Ramaswamy (1979)), only recently have researchers begun to formally analyze the theoretical basis for doing such sorts. Lo and MacKinlay (1990) point out that if the sort is based on either a variable that is only known to be empirically correlated with returns or a variable measured within the sample, the test will contain a data-snooping bias. Liang (2000) argues that even when the sort is based on a variable estimated using prior data, measurement error in this variable can lead to false conclusions. In this paper we will focus on a different variation of the sorting technique analyzed in those papers.

The empirical procedure that is the subject of this paper closely resembles the Black et al. (1972) grouping procedure. There is, however, one crucial distinction. Like the standard procedure, assets are sorted into groups using some criterion related to asset returns. However, rather than forming portfolios out of the groups, the tests are run *within* the groups. It is shown that this empirical procedure biases the results in favor of rejecting whatever asset pricing model is being tested. In particular, we show the following:

1. The explanatory power of the model will always be smaller within a group than in the whole sample.
2. By picking enough groups to sort into, a researcher can destroy the within-group explanatory power of even an economically correct asset pricing model.

* University of California, Berkeley, and NBER. The ideas in this paper were first presented by the author as formal discussions at both the 1996 and 1997 meetings of the American Finance Association. The author thanks Kent Daniel, Wayne Ferson, Ken French, Cam Harvey, Jens Jackwerth, Raymond Kan, Bing Liang, Tim Loughran, Paul Malatesta, Sheridan Titman, the editor, and three anonymous referees for their comments on earlier drafts of this paper.

To understand the intuition behind these results, consider the following thought experiment. Assume that a researcher has uncovered a variable, γ, that is cross-sectionally correlated with returns. He postulates that he has uncovered an anomaly—that is, he claims that γ is a better predictor of return than a particular risk-based asset pricing model, hereafter the "SOS" model. The reason is that he has sorted stocks into groups based on γ and shown that within any of these groups the SOS asset pricing model has little or no explanatory power.

When stocks are first sorted into groups based on a variable like γ that is known to be correlated with returns, the average return of each group will reflect this correlation. Consequently, the procedure ensures a high between-group variation in expected return. A standard result of ANOVA hypothesis testing is that the total variance of the sample is the sum of the between-group and within-group variance. Thus, the high between-group variation in expected return implies a low within-group variation, so the cross-sectional variation of expected returns is lower within the groups than in the whole sample.

Now, no economic model makes exact predictions. So the expected return predicted by the SOS model will not equal the true expected return. Assume the SOS model is nevertheless a good description of reality so that the SOS model's error is unbiased. Then the prediction of the model will be the true expected return plus mean zero noise. If the model error is distributed independently of the sort variable, the sort will not affect the cross-sectional distribution of the noise. The implication is that cross-sectional variance of the model error will be no different within the groups than in the full sample. Thus, the within-group ratio of the cross-sectional variance of expected returns to the variance of the model error must be lower in the groups than in the full sample. In short, the model's signal-to-noise ratio is lower within the groups than the full sample. Indeed, no matter how small the model error is, by sorting into enough groups, the variation in expected returns can be reduced to the point that the error in the SOS model completely swamps the explanatory power of the model. So long as the model cannot predict the expected return of each stock exactly, it is possible for our researcher to demonstrate that the SOS model has no explanatory power within the γ sorted groups, even when the SOS model does in fact provide economically useful predictions.

In a sense, one can think of this point as the converse of a point made originally in Black et al. (1972). There, the authors argue that to provide a powerful test of the CAPM it is essential that the stocks are grouped into portfolios that have large cross-sectional differences in expected returns. In this context, large cross-sectional differences in portfolio expected returns imply that the cross-sectional variation in the expected return of the stocks that make up the portfolios is small. Thus, if instead of proceeding as Black et al. do, you decide to test the CAPM within the stocks that make up the portfolios rather than on the portfolios themselves, the opposite result holds. To increase the power of this test, the cross-sectional variation within each

portfolio group must be maximized, which implies the cross-sectional differences in the portfolio expected returns must be minimized. Perhaps the simplest way to do this is to just not group the stocks and use the full sample.

Although the particular sorting procedure that we analyze is not new,[1] one motivation for studying it derives from the fact that recently a number of studies employing this procedure have produced rather disturbing results. For instance, Daniel and Titman (1997) sort stocks into fractiles based on their characteristics (book-to-market ratios and market values) and show that the factors identified by Fama and French (1993) cannot explain the within-fractile variation in realized return. Based on this result, they conclude that asset returns are likely generated by a characteristics-based asset pricing model. If asset returns are indeed explained by a characteristics-based model, then this would upset some of the foundations of asset pricing theory.

The rest of the paper is organized as follows. In Section I we illustrate our results in the context of a simple example. We take a single-period economy in which the CAPM holds exactly and we sort stocks into groups based on the book-to-market ratio. We then derive the theoretical results of a within-group cross-sectional regression test and show that the model has little or no explanatory power in all but the extreme groups. That is, even though the CAPM holds equally well for all stocks, the test leaves the impression that the model only holds in the highest and lowest groups. Section II then derives the main theoretical results. The implications of these results are considered in Section III. In Section IV we illustrate some of these implications in the context of one recent study—Daniel and Titman (1997). Section V concludes the paper. All proofs are in the Appendix.

I. An Illustrative Example

Before we derive the main theoretical results in this paper, it is useful to illustrate their potential importance in the context of a simple example. Consider a one-period world in which the CAPM holds exactly. If the total return of stock i is R_i, then, since the CAPM holds exactly,

$$E[R_i] = r + \beta_i(E[R_m] - r), \tag{1}$$

where r is the risk-free rate, β_i is the beta of stock i, and R_m is the total return of the market portfolio.

Next, construct a cross-sectional test of this model. Of course, neither $E[R_i]$ nor β_i is directly observable. Instead of the expected return, the realized, or measured, return, $\hat{R}_i$, where

$$\hat{R}_i = E[R_i] + \xi_i, \tag{2}$$

[1]Litzenberger and Ramaswamy (1980), for instance, employ this procedure.

is observed. We assume that since ξ_i is the deviation from the expected return, it has mean zero and variance ω^2, so that the distribution of ξ_i does not differ across stocks. Instead of the true beta, β_i, the measured beta, $\hat{\beta}_i$, where

$$\hat{\beta}_i = \beta_i + \epsilon_i, \tag{3}$$

is observed. Since ϵ_i reflects measurement error, we assume that it is Normal$[0,\theta]$ and independent of everything else in the economy.[2]

All that is left to specify is the cross-sectional distribution of firms in the economy. The cross-sectional distribution of a firm's expected cash flow turns out not to affect the results, so we let it be arbitrary. We assume that the cross-sectional distribution of β_i is Normal$[1,\sigma]$, so that the market portfolio has a beta of one. For now, assume that the cross-sectional covariance between beta and ξ_i is zero, although in the next section we demonstrate in a more general setting that this must be the case. To abstract away from the problems associated with small sample sizes, we assume that infinitely many stocks exist. To emphasize that only a small amount of measurement error is required to deliver the result, we take $\theta^2 = 0.025\sigma^2$, so the cross-sectional variance of $\hat{\beta}_i$ is a mere 2.5 percent larger than the cross-sectional variance of β_i.

First, consider running a cross-sectional regression of the realized risk premium, $\hat{R}_i - r$, on the measured beta, $\hat{\beta}_i$, in the full sample. The coefficient in this regression is

$$\begin{aligned}
\frac{\operatorname{cov}(\hat{R}_i - r, \hat{\beta}_i)}{\operatorname{var}(\hat{\beta}_i)} &= \frac{\operatorname{cov}(E[R_i] + \xi_i - r, \beta_i + \epsilon_i)}{\operatorname{var}(\beta_i + \epsilon_i)} \\
&= \frac{\operatorname{cov}(E[R_i] - r, \beta_i)}{\operatorname{var}(\beta_i) + \operatorname{var}(\epsilon_i)} \\
&= \frac{\operatorname{cov}(\beta_i(E[R_m] - r), \beta_i)}{\operatorname{var}(\beta_i) + \operatorname{var}(\epsilon_i)} \\
&= (E[R_m] - r)\frac{\sigma^2}{\sigma^2 + \theta^2} \\
&= (E[R_m] - r)\frac{\sigma^2}{\sigma^2 + 0.025\sigma^2} = 0.976(E[R_m] - r),
\end{aligned} \tag{4}$$

[2] Since the value-weighted sum of the measurement error must be zero, the independence assumption cannot be satisfied in any finite economy. However, it is theoretically possible to make this assumption in an economy with infinitely many assets because the variance of the value-weighted sum of the measurement error converges to zero as the number of assets goes to infinity.

where the second line follows from the independence assumption, the third from equation (1), and the fourth from the size of the sample. Because the attenuation bias is relatively small, the coefficient is close to the theoretical value of $E[R_m] - r$. Similar logic shows that the intercept is $0.024(E[R_m] - r)$, which again is close to the theoretical value of 0. Finally, the R^2 coefficient is

$$\begin{aligned} & 0.976^2(E[R_m] - r)^2 \frac{\text{var}(\hat{\beta}_i) + 1}{\text{var}(\hat{R}_i - r) + (E[R_m] - r)^2} \\ &= 0.976^2 \frac{\text{var}(\hat{\beta}_i) + 1}{\text{var}(\beta_i) + \text{var}(\xi_i) + 1} \\ &= 0.976^2 \frac{\sigma^2 + \theta^2 + 1}{\sigma^2 + \omega^2 + 1}. \end{aligned} \tag{5}$$

If we assume that the expected return can be measured with at least the degree of accuracy of β—that is, if $\omega \leq \theta$—then we have that $R^2 \leq 0.976^2 = 95\%$. Given that the model holds exactly and that the amount of measurement error is minimal, such results are not surprising.

Next we run a similar test, this time with the data first sorted into N fractiles based on the firms' book-to-market ratios. That is, each fractile contains equally many stocks and the book-to-market ratios within each fractile are strictly larger (smaller) than the book-to-market ratios of the next lower (higher) fractile. Then within each fractile the same test is undertaken.

To keep the example simple we assume that the terminal cash flow of each firm is proportional to its scale. Consequently, the book value of assets, B_i, is assumed to be proportional to the terminal dividend or expected cash flow $E[C_i]$ of each stock,

$$E[C_i] = KB_i, \tag{6}$$

where $K > 0$, the constant of proportionality, is assumed to be the same for all firms. This assumption implies that the book-to-market ratio is correlated with expected returns. Since the CAPM holds exactly, the market value or price, p_i, is given by

$$p_i = \frac{E[C_i]}{E[R_i]} = \frac{E[C_i]}{r + \beta_i(E[R_m] - r)}. \tag{7}$$

To compute the results of these regressions we need to know the cross-sectional variance of beta conditional on being in a particular fractile. First define $U(j)$ as the largest book-to-market ratio in the jth fractile. Then, if σ_j is defined to be the cross-sectional variance of the true beta conditional on being in fractile j,

$$\begin{aligned}\sigma_j^2 &\equiv \operatorname{var}\left(\beta_i \middle| U(j-1) < \frac{B_i}{p_i} \leq U(j)\right)\\ &= \operatorname{var}\left(\beta_i \middle| U(j-1) < \frac{E[C_i]}{\dfrac{KE[C_i]}{r + \beta_i(E[R_m] - r)}} \leq U(j)\right)\\ &= \operatorname{var}(\beta_i | U^*(j-1) < \beta_i \leq U^*(j)),\end{aligned} \tag{8}$$

where $U^*(j) = (KU(j) - r)/(E[R_m] - r)$ and $U(0) \equiv -\infty$. Let $\Psi(\cdot)$ be defined as the standard normal distribution function. From equation (8), it follows that

$$\begin{aligned}\sigma_j^2 = \sigma^2 &\int_{\frac{U^*(j-1)-1}{\sigma}}^{\frac{U^*(j)-1}{\sigma}} \frac{\beta^2}{\Psi\left(\dfrac{U^*(j)-1}{\sigma}\right) - \Psi\left(\dfrac{U^*(j-1)-1}{\sigma}\right)} d\Psi(\beta)\\ &- \sigma^2 \left(\int_{\frac{U^*(j-1)-1}{\sigma}}^{\frac{U^*(j)-1}{\sigma}} \frac{\beta}{\Psi\left(\dfrac{U^*(j)-1}{\sigma}\right) - \Psi\left(\dfrac{U^*(j-1)-1}{\sigma}\right)} d\Psi(\beta) \right)^2\\ &= \sigma^2 g(j),\end{aligned} \tag{9}$$

where $g(\cdot)$, defined implicitly above, is a function only of the fractile j and can be calculated from the standard normal alone. One can think of $g(\cdot)$ as providing a measure of the difference between each within-category variance and the full-sample variance. Using equation (9) and following the same logic as before, the cross-sectional regression coefficient within the jth fractile is

$$\begin{aligned}(E[R_m] - r)\frac{\sigma_j^2}{\sigma_j^2 + 0.025\sigma^2} &= (E[R_m] - r)\frac{\sigma^2 g(j)}{\sigma^2 g(j) + 0.025\sigma^2}\\ &= (E[R_m] - r)\frac{g(j)}{g(j) + 0.025}.\end{aligned} \tag{10}$$

Table I gives the value of $g^*(j) \equiv g(j)/(g(j) + 0.025)$ in each fractile for three tests that differ only in the number of total fractiles stocks are sorted into. The following results are apparent in all three tests. If the $g^*(\cdot)$'s in the table are used to compute the regression coefficient and intercept (from equation (10)), then (i) all the within-group coefficients (intercepts) are lower (higher) than the coefficient obtained in the full sample and (ii) in most cases the coefficients (intercepts) are well below (above) any reasonable cutoff for concluding that they provide support for the model. In all three tests only two fractiles produce coefficients that are even remotely close to the

Table I

Within-Fractile Regression Results

This table contains the within-fractile value of a measure of how much lower each within-sample variance is than the full-sample variance, $g^*(j) \equiv g(j)/(g(j) + 0.025)$, and the within-fractile value R^2 for three different partitions of the data into 10 (Panel A), 20 (Panel B), and 100 (Panel C) fractiles. To get the within-fractile regression coefficient, the number in the first row of each panel, $g^*(j)$, should be multiplied by the market risk premium, $E[R_m] - r$. Multiplying the number in the second row of each panel, $1 - g^*(j)$, by $E[R_m] - r$ provides the intercept. The third row is the R^2 coefficient, $(g^*(j))^2$, when the measurement error in realized returns is of the same order of magnitude as the measurement error in beta (i.e., $\omega = \theta$). Except for the 10-fractile partition, the table lists these values for a subset of the fractiles. The listed fractiles are the column headings in each panel. The final column in the table gives the average of each variable over all fractiles.

Panel A: 10-Fractile Partition

Fractile	1	2	3	4	5	6	7	8	9	10	Average
$g^*(j)$	0.87	0.39	0.25	0.20	0.18	0.18	0.20	0.25	0.39	0.87	0.38
$1 - g^*(j)$	0.13	0.61	0.75	0.80	0.82	0.82	0.80	0.75	0.61	0.13	0.62
R^2	0.76	0.15	0.063	0.039	0.031	0.031	0.039	0.063	0.15	0.76	0.208

Panel B: 20-Fractile Partition

Fractile	1	2	6	8	10	12	14	16	19	20	Average
$g^*(j)$	0.85	0.3	0.07	0.055	0.050	0.052	0.061	0.085	0.3	0.85	0.18
$1 - g^*(j)$	0.15	0.7	0.93	0.945	0.950	0.948	0.939	0.915	0.7	0.15	0.82
R^2	0.72	0.091	0.0049	0.0030	0.0025	0.0027	0.0037	0.0072	0.091	0.72	0.087

Panel C: 100-Fractile Partition

Fractile	1	2	30	40	50	60	70	80	99	100	Average
$g^*(j)$	0.79	0.2	0.0028	0.0022	0.0021	0.0022	0.0027	0.0041	0.2	0.79	0.028
$1 - g^*(j)$	0.21	0.8	0.9972	0.9978	0.9979	0.9978	0.9973	0.9956	0.8	0.21	0.97
R^2	0.63	0.038	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.038	0.63	0.014

theoretical value of one. In all cases these two fractiles are the extremes—the lowest and the highest, presumably because these fractiles are the least affected by the truncation bias induced by the sort.

Similar inferences can be made about the explanatory power of the model. Consider the 10-fractile grouping. Under the assumption that $\omega = \theta$, which provides an R^2 of 95 percent in the full sample, the R^2 coefficient is given by $(g^*(\cdot))^2$. With the exception of the extreme groups, in all other groups the R^2 coefficients are below 16 percent and most are below seven percent. Indeed, the lowest R^2 is a mere 3.1 percent. When the number of groups is increased to 20 or more, the vast majority of R^2 coefficients essentially shrink to zero. A naive interpretation of these results would be that if the CAPM holds at all, it only holds among stocks with extreme book-to-market ratios. In fact, the CAPM holds equally well for all stocks by assumption. Although this is merely an example, it does illustrate the main point of the paper. By first sorting on variables such as book-to-market, it is quite possible for the model to appear to have little or no explanatory power even if it holds exactly.

In this example, by sorting stocks by a variable such as book-to-market that is known to be correlated with expected returns, the researcher essentially does an expected return sort. Thus, when groups are formed based on this sort, the within-group cross-sectional variation in expected return is lower than in the whole sample. Since the model error (measurement error in beta) is independent, it is unaffected by the sort; that is, the distribution of measurement error in beta is no different within a group than in the whole sample. Thus, the signal-to-noise ratio—the ratio of the variance of expected returns to the variance of the model error—is lower within a group. Since the within-group cross-sectional variation in expected returns is decreasing in the number of groups, by sorting into enough groups the relative importance of the model error can be increased to the point that it accounts for almost all the variation in the model's prediction of the expected return.

It is worth noting that if, instead of testing the model within the groups, the Black et al. (1972) procedure is followed and equally weighted portfolios are formed out of the stocks within a group, the returns on these portfolios would provide an *exact* fit of the CAPM. This follows because, under our assumption that infinitely many stocks exist, the betas of the equally weighted portfolios would not contain any measurement error and so no model error would exist. In the context of this example, by undertaking one simple variation of the Black et al. sorting procedure, the empiricist greatly biases the test in favor of rejecting the model under consideration.

Although one might justifiably argue that this example is extreme, it does illustrate concisely the main intuition of the paper. Our objective in the next section is to establish conditions under which these results will generally be true.

II. Theory

We take as a primitive an economy in which the observed or measured return at time t of every firm i is $\hat{R}_{it}$, where

$$\hat{R}_{it+1} = E_t[R_{it+1}] + \xi_{it+1}. \tag{11}$$

ξ_{it+1} is assumed to have mean zero. We again assume that a countably infinite number of stocks exist. Given the information set at time t, let $F_t(\cdot)$ be the cross-sectional distribution of $E_t[R_{it+1}]$. Let the mean and variance of this distribution be denoted μ_t and σ_t^2 and assume it has bounded support. The cross-sectional variance of ξ_{it+1} is denoted ω_t^2. Note that since ξ_{it+1} is an expectational error it is uncorrelated to anything in the information set at time t. Since $E_t[R_{it+1}]$ is in this information set, the cross-sectional correlation between ξ_{it+1} and $E_t[R_{it+1}]$ is zero.

We further assume that an empiricist chooses to undertake a test of an asset pricing model in this economy. Denote $\bar{R}_{it+1}$ as the prediction of the model of $E_t[R_{it+1}]$, given the information set at time t. Of course, like any economic model it is not *exact*—$\bar{R}_{it+1} \neq E_t[R_{it+1}]$—the expected return calculated using the model does not exactly equal the actual expected return. The reason for this difference is not relevant—it could result from model misspecification, parameter observability issues (e.g., an error-in-variables problem), or microstructure effects. The difference simply reflects the fact that no economic model is capable of making predictions that are perfectly accurate. Even the predictions of the true asset pricing model are affected by errors in measuring the model's parameters. Therefore, let

$$\bar{R}_{it+1} = E_t[R_{it+1}] + \epsilon_{it+1}, \tag{12}$$

where ϵ_{it+1} is distributed independently of everything else in the economy and $E_t[\epsilon_{it+1}] = 0$. Let the cross-sectional variance of ϵ_{it+1}, given the information set at time t, be denoted θ_t^2. Under these assumptions and given the information set at time t, the cross-sectional distribution of ϵ_{it+1} is independent of the cross-sectional distribution of $E_t[R_{it+1}]$ as well as of ξ_{it+1}. Thus, the nature of the model error is such that it is unpredictable given any conditioning information—the pricing model is good enough so that it does not systematically misprice stocks.

The advantage of focusing on the prediction of the asset pricing model, $\bar{R}_{it+1}$, rather than on the model itself, is that this allows a greater degree of generality because we do not have to restrict the functional form of the model itself. By definition, any asset pricing model must contain a prediction for the expected return. Therefore, the results in this paper can, in principle, be applied to a test of any of the standard linear models as well as other, possibly nonlinear, asset pricing models not yet derived.

One procedure for testing the asset pricing model is to cross-sectionally regress $\hat{R}_{it+1}$ (the observed return) onto $\bar{R}_{it+1}$ (the prediction of the model) at each point in time.[3] Given the information set at time t, the coefficient of this regression is

$$\begin{aligned}\frac{\text{cov}(\bar{R}_{it+1},\hat{R}_{it+1})}{\text{var}(\bar{R}_{it+1})} &= \frac{\text{cov}(E_t[R_{it+1}]+\xi_{it+1},E_t[R_{it+1}]+\epsilon_{it+1})}{\text{var}(E_t[R_{it+1}]+\epsilon_{it+1})}\\ &= \frac{\sigma_t^2}{\sigma_t^2+\theta_t^2}\\ &= \frac{1}{1+\left(\frac{\theta_t}{\sigma_t}\right)^2},\end{aligned} \tag{13}$$

where the second line follows from the independence of ϵ_{it+1} and the fact that ξ_{it+1} and $E_t[R_{it+1}]$ are cross-sectionally uncorrelated. When the model holds exactly ($\theta_t = 0$) the coefficient is one. When the model does not hold exactly the coefficient is always less than one, and this attenuation is an increasing function of the difference between the model's prediction of the expected return and the actual expected return. For expositional simplicity we henceforth drop the explicit dependence on time (except in cases where this would lead to ambiguity).

An important feature of the example in the previous section is that the sort is undertaken using a variable, namely the book-to-market ratio, that is known a priori to produce groups with different average returns. This is the key assumption required to deliver the result. Thus, instead of running the above cross-sectional regression in the whole sample, at each point in time the stocks are first grouped into n nonempty fractiles, denoted $j(n) = 1,\ldots,n$, for which it is known, a priori, that the expected return in each fractile differs.[4] Note that the exact procedure followed to construct the fractiles is immaterial. For example, the groups can be produced by sorting by a variable that is known to be correlated (though not necessarily perfectly)[5] with returns.

Let the conditional cross-sectional distribution of expected return in fractile $j(n)$ be denoted $F^{j(n)}(E[R_i]) \equiv F(E[R_i]|i \in j(n))$ with $Q_{j(n)} \equiv \int_{i\in j(n)} dF$. Then define $E_{j(n)}$ as the expected return (at time t) of all stocks within fractile $j(n)$; that is,

$$E_{j(n)} \equiv \int_{i\in j(n)} E[R_i]\,dF^{j(n)}, \tag{14}$$

[3] Almost any cross-sectional test of an asset pricing model can be fitted into this framework. For instance, this procedure is equivalent to a cross-sectional test of a single-factor model because the factor risk premium is common to all stocks. Thus, regressing on the factor beta is equivalent to regression on the beta times the risk premium. If the model being tested is a multifactor model with factor risk premia λ_k, then $\bar{R}_{it} = r + \sum_{k=1}^{K} b_{ki}\lambda_k$.

[4] Thus $j(n)$ is the jth group in a partition consisting of n groups. The reason for explicitly writing the group number as a function of the number of groups will become clear shortly.

[5] Perfect correlation is a necessary but not a sufficient condition.

and define $\sigma^2_{j(n)}$ as the cross-sectional variance of the expected return (at time t) of all stocks within fractile $j(n)$; that is,

$$\sigma^2_{j(n)} \equiv \int_{i \in j(n)} (E[R_i])^2 \, dF^{j(n)} - E^2_{j(n)}. \tag{15}$$

The (weighted) average cross-sectional variation of the within-fractile expected return is strictly smaller than the full-sample variation. Furthermore, as the number of fractiles goes to infinity, the average within-fractile variation goes to zero. The following lemma proves this.

LEMMA 1. *Given any* $\theta > 0$,

1. *For any* $n > 1$, $\sum_{j(n)=1}^{n} Q_{j(n)} \sigma^2_{j(n)} < \sigma^2$.
2. *For any* $0 < \delta < \sigma^2$, *there exists an* N *such that for all* $n > N$

$$\sum_{j(n)=1}^{n} Q_{j(n)} \sigma^2_{j(n)} < \delta.$$

Now consider what might happen if the empiricist proceeds to run the cross-sectional regression test within each fractile. Using the same logic that provided equation (13), the coefficient of the test in the jth fractile is

$$\frac{\sigma^2_{j(n)}}{\sigma^2_{j(n)} + \theta^2} = \frac{1}{1 + \left(\dfrac{\theta}{\sigma_{j(n)}}\right)^2}. \tag{16}$$

An implication of the above lemma is that a grouping of stocks always exists in which every within-fractile coefficient (i.e., equation (16)) is less than the full-sample coefficient (i.e., equation (13)). Furthermore, by choosing a large enough number of fractiles, the empiricist can make every within-fractile coefficient as close to zero as he wants.

Before we prove these facts we need some more definitions. Let a *grouping* be any partition of the dataset into $n > 1$ fractiles for which the average expected return of each fractile is distinct. A *series of groupings* is then any ordered set of groupings in which the number of fractiles, n, in each grouping increases monotonically and each fractile in the n^{th} grouping is a subset of a fractile in the $n - 1$th grouping.

PROPOSITION 1. *Consider the set of all possible groupings with* n *fractiles. Then for any* $\theta > 0$, *there is a nonempty subset of this set such that for all groupings in the subset*

$$\frac{1}{1 + \left(\dfrac{\theta}{\sigma_{j(n)}}\right)^2} < \frac{1}{1 + \left(\dfrac{\theta}{\sigma}\right)^2}, \qquad 1 \le j(n) \le n.$$

The proof of this proposition shows that so long as the within-fractile variation does not differ by too much between fractiles, the regression coefficient within every fractile will be lower than the coefficient for the whole sample. It is worth emphasizing the implications of this result. The result requires the existence of some variation in expected returns between fractiles. Satisfying this condition requires little more than identifying a variable known to be correlated to expected returns and using this variable to form the fractiles. The magnitude of this correlation need not be very large, in particular, it need not be one. The implication is that even for a variable that is only weakly related to expected returns, the empirical procedure will be subject to the bias identified in the proposition if such a variable is used to form the fractiles.

PROPOSITION 2. *Consider the set of all possible series of groupings. Take any* $\theta > 0$. *Then for any* $0 < \eta < 1/(1 + (\theta/\sigma)^2)$, *there exists a nonempty subset of this set and an* N *such that for all* $n > N$

$$\frac{1}{1 + \left(\frac{\theta}{\sigma_{j(n)}}\right)^2} < \eta, \qquad 1 \leq j(n) \leq n.$$

The proof of the proposition shows that by picking enough fractiles (with sufficiently similar within-fractile variation), the coefficients in all fractiles can be made arbitrarily close to zero. The R^2 coefficient of each within-group regression is

$$\begin{aligned}\left(\frac{1}{1 + \left(\frac{\theta}{\sigma_{j(n)}}\right)^2}\right)^2 \left(\frac{\text{var}(\bar{R}_i) + \mu^2}{\text{var}(\hat{R}_i) + \mu^2}\right) &= \left(\frac{1}{1 + \left(\frac{\theta}{\sigma_{j(n)}}\right)^2}\right)^2 \left(\frac{\theta^2 + \sigma_{j(n)}^2 + \mu^2}{\omega^2 + \sigma_{j(n)}^2 + \mu^2}\right) \\ &\leq \left(\frac{1}{1 + \left(\frac{\theta}{\sigma_{j(n)}}\right)^2}\right)^2 \left(\frac{\theta^2 + \sigma_{j(n)}^2}{\sigma_{j(n)}^2}\right) \qquad (17) \\ &= \frac{1}{1 + \left(\frac{\theta}{\sigma_{j(n)}}\right)^2}.\end{aligned}$$

Consequently, the following corollary follows immediately from Proposition 2.

COROLLARY 1. *Consider the set of all possible series of groupings. Take any* $\theta > 0$. *Then for any* $0 < \eta < 1/(1 + (\theta/\sigma)^2)$, *there exists a nonempty subset of this set and an* N *such that for all* $n > N$ *every within-fractile* R^2 *coefficient will be strictly less than* η.

In words, regardless of how well the model fits the full sample, the empiricist can ensure that the within-fractile fit is arbitrarily bad by simply picking a large enough number of fractiles and making sure the within-fractile variance does not differ by too much between fractiles. The bottom line is that if the empiricist proceeds naively by judging the explanatory power of the model within any fractile using the same criteria as he does to judge the explanatory power in the full sample, then simply by judiciously picking enough fractiles he can reduce the model's explanatory power to zero.

III. Implications

The results derived in the previous section rely on the fact that the average within-fractile variance of expected returns is always less than the full-sample variance. Equation (A4) in the proof of the lemma formalizes this notion. It also shows that, regardless of the number of fractiles, the weighted-average variance of all the fractiles is always lower than the full-sample variance. The amount of this "loss" is given by the strictly positive term $\sum_{j(n)=1}^{n} Q_{j(n)}(E^*_{j(n)})^2$, or what is generally know as the between-fractile variation. This variation is a measure of how much "worse" the asset pricing model will do within the fractiles. The implication is that the greater the cross-sectional differences in expected returns across fractiles are, the worse the within-fractile performance of the asset pricing model will be. As was pointed out in the introduction, one can think of this point as the converse of a point made originally in Black et al. (1972). There, the authors argued that to provide a powerful test of the CAPM it is essential that the stocks are grouped into *portfolios* that have large cross-sectional differences in expected returns (see Huang and Litzenberger (1988, Chap. 10), for an excellent discussion of this point). By equation (A4), large cross-sectional differences in portfolio expected returns imply that the cross-sectional variation in the expected return of the stocks that make up the portfolios must be small. Thus, if you decide to test the CAPM within the stocks that make up the portfolios rather than on the portfolios themselves, you should do the opposite of what Black et al. recommend. To increase the power of this test, the cross-sectional variation within each portfolio group must be maximized, which implies the cross-sectional differences in the portfolio expected returns must be minimized.

Lemma 1 shows that so long as there is any between-group variance in expected returns, the average within-group variance will be lower than the variance of expected returns in the full sample. Consequently, the most effective way to maximize the within-group variance is to have only one group—in other words, do not sort at all and use the full sample. However, there might be valid reasons for not using the whole sample. One frequent justification for this kind of sorting is to determine whether an asset pricing model can explain a known characteristic of the data, for instance, a positive

correlation between a particular variable and return. An implication of the results is that the criteria for the sorts are important in setting up this kind of empirical test.

Suppose a researcher wanted to test whether a particular asset pricing model, say the SOS model, can explain an empirically observed cross-sectional relation between a variable, denoted γ, and average return. She therefore takes as her Null hypothesis the hypothesis that the SOS asset pricing model is correctly specified and holds. A necessary condition for this hypothesis to hold is that the SOS model explain the cross-sectional relation between γ and average return. The Alternative hypothesis is that the SOS model does not hold. The key point to note is that γ is related to average returns under *both* hypotheses. Consequently, if stocks are first sorted by γ, the ability of the SOS model to discriminate within the groups will be diminished under both hypotheses. If the implications of the propositions are ignored, then this could lead to a false rejection of the Null. If, however, stocks are first sorted by the expected return predicted by the SOS model, the within-group relation between γ and average returns will only be diminished under the Null. This follows because under the Alternative the model cannot differentiate stocks with different expected returns and so there is no reason to expect between-group variation in expected returns. Thus the correct way to implement a test of this Null is to first sort by the prediction of the SOS model and then see whether, within the groups, γ is still related to expected return.

There are at least two other reasons why researchers group data. First, grouping data can significantly reduce the computational burden associated with the size of financial data sets. However, in recent years advances in computational speed have significantly relaxed this constraint, so more recently authors have argued for techniques that do not require grouping (see, e.g., Kim (1996)).

The other reason for grouping the data concerns the number of degrees of freedom in cross-sectional studies that require estimating the variance-covariance matrix of returns. Given a typical sample of 2000 stocks, this matrix has more than 2 million elements. With only 70 years of data, there is an obvious specification problem. Since this problem is unlikely to be solved in the foreseeable future, it appears that grouping the data will always be an integral part of these studies. It is therefore worth considering what implications the results in this paper have for those studies.

Our results rely on the assumption that there exists between-group variation in expected returns. Such variation can arise because the researcher chose to construct the groups in a sample in which other studies already detected this variation. It can also arise because under one or more of the hypotheses being tested, the between-group variation is theoretically predicted. In this case, unless the researcher explicitly accounts for the effect of this variation, the power of the test against this hypothesis will be reduced.

In the next section we briefly illustrate the implications of our results in the context of a study that uses the sorting procedure examined in this paper.

Table II

Mean Excess Returns

This table is based on Daniel and Titman's (1997) Table III, and presents the mean excess returns of the 45 portfolios formed on the basis of market value (SZ), book-to-market (BM), and the estimated factor loadings on the HML portfolio (i.e., the betas on the HML factor) for the period from July 1973 through December of 1993. Each of the five factor-loading portfolio columns provides the monthly excess returns of portfolios of stocks that are ranked in a particular quintile with respect to the HML factor loading (with column 1 being the lowest and column 5 being the highest). The firm size and book-to-market rankings of the stocks in each of the portfolios are specified in the nine rows. For example, the row 1, portfolio 1 entry (0.148) is the mean excess return of a value-weighted portfolio of the stocks that have the largest market value, the lowest book-to-market, and the lowest expected loading on the HML factor. The last two columns of this table do not appear in the corresponding table in Daniel and Titman. The second to last column is just the average of the returns reported in the table of the five factor-loading portfolios. The last column is the difference between the sum of the reported returns for portfolios 4 and 5 and the sum of the reported returns for portfolios 1 and 2.

Char. Port.		Factor-Loading Portfolio						
BM	SZ	1	2	3	4	5	Average	Portfolio (4 + 5) − (1 + 2)
1	3	0.148	0.287	0.396	0.400	0.830	0.4122	0.795
2	3	0.645	0.497	0.615	0.572	0.718	0.6094	0.148
1	1	0.202	0.833	0.902	0.731	0.504	0.6344	0.200
1	2	0.711	0.607	0.776	0.872	0.710	0.7352	0.264
3	3	0.736	0.933	0.571	0.843	0.961	0.8088	0.135
2	2	0.847	0.957	0.997	0.873	0.724	0.8796	−0.207
2	1	1.036	0.964	1.014	1.162	0.862	1.0076	0.024
3	2	1.122	1.166	1.168	1.080	0.955	1.0982	−0.253
3	1	1.211	1.112	1.174	1.265	0.994	1.1512	−0.064
Average		0.740	0.817	0.846	0.866	0.806		

IV. An Application: Daniel and Titman

Based on their results, Daniel and Titman (1997) argue that a characteristics-based model does a better job of explaining the cross section of asset return than the Fama–French factors do. In this section we will illustrate the implications of our propositions on this conclusion.

We will concentrate on the results contained in Daniel and Titman's (1997) Table III, which is reproduced here (with additional material) as Table II. To generate their table, Daniel and Titman first sort stocks into nine different groups based on the market value and book-to-market ratio, hereafter BM/SZ groups.[6] The authors knew, a priori, that such a sort will produce groups in which the expected returns of the groups will differ because they use a sam-

[6] They sort stocks by market value (book-to-market) and establish 33.3 percent and 66.6 percent breakpoints. Based on these breakpoints, the authors define nine possible groups and sort stocks into those nine groups based on their market value and book-to-market ratios.

ple for which it is well known from past empirical studies (in particular, the study that motivated the research in this area—Fama and French (1992)) that these variables are highly correlated with realized returns.

The authors next test whether a factor model can distinguish stocks within each BM/SZ group. The factors they use are the three factors previously identified empirically by Fama and French (1993). They take each stock's loading on one of these factors (HML) (i.e., the beta of each stock on the HML factor), and within each BM/SZ group they sort stocks into five portfolios based on this loading. They then calculate the return on these portfolios and, based on the fact that within each BM/SZ group there is almost no systematic difference between the returns of these portfolios, they conclude that once firm "characteristics" are controlled for, there is only a weak relation at best between factor loadings and returns.

Their results are reproduced here in Table II with two additions. Each row in the table shows the return of the five portfolios sorted on HML factor loading for each BM/SZ group. We have added two columns, the average return across the factor-loading portfolios (second to last column) and the difference between portfolios 4 plus 5 and 1 plus 2 (last column). The first addition gives some idea of what the average return in each BM/SZ group is and the second speaks to the ability of the factors to differentiate returns. We have also rearranged the rows so they are now ordered increasing in realized return.

There is one thing, in particular, that is worth noting about the table. The difference between the average realized return of the lowest and highest BM/SZ group is remarkably large (0.74 percent/month). On an annual basis this difference works out to be about 9.24 percent. To the extent that the large realized differences in average returns reflect large differences in expected returns, this sorting procedure reduces the cross-sectional variation in expected returns within each group. Furthermore, even if these factors completely explain asset returns, since the factor loadings are not observed directly, they are measured with error. This error alone is enough to ensure the condition that the model error is nonzero; that is, $\theta > 0$ is satisfied. There is, however, a potentially more important source of model error. Since stocks are ranked on only one of three factors, the contribution of the other two factors is "missing." Thus the "missing" factors are additional sources of model error—a high beta on the factor might not be associated with a high return because the other betas on the factors not ranked are low. Under their Null, all three factors are hypothesized to explain returns, so by leaving out two factors the authors are implicitly ensuring that θ is large under the Null. The reduced within-group cross-sectional variation in expected returns coupled with the model error explains why these authors fail to find a discernible relation between the factor loading and returns in the data set in which these factors were identified.

It is important to understand that not all of the implications of the Daniel and Titman study rely on the results of the above sorts. For instance, they undertake a series of Black et al. (1972) type tests of the Null hypothesis

that the Fama–French factors explain returns. Since none of these tests are run within the BM/SZ groups, they are not subject to the bias discussed in this paper. Under their Null hypothesis, the regression intercept should be zero, yet they find that three of the nine intercepts have t-statistics above two. Although they do not conduct a joint test, at first glance anyway, it is hard to disagree with their conclusion that the intercept is nonzero. Taking this result at face value, their inference—that we can reject the Null hypothesis that the Fama–French factors correctly price assets—is justified. This result is reminiscent of earlier studies that have used other factors to derive the same result. However, Daniel and Titman claim an additional result in their paper—that their characteristics-based model does better than the Fama–French factors. Results of the sort analyzed above provide the evidence in their paper that supports this claim. Without this evidence it is hard to justify this more controversial result.

V. Conclusion

The results in this paper imply that conclusions of empirical studies that first sort the data into groups, based on a variable known a priori to be correlated with returns, and then run tests within the groups should be questioned. It is shown that such a procedure biases the test in favor of rejecting the asset pricing model under consideration. We demonstrate that simply by sorting into enough groups, it is possible for a researcher to reduce the explanatory power of even an economically correct asset pricing model to zero.

The question of how sorting techniques affect the statistical inferences of the studies that employ them is important because, given the realities of the data, it is likely that financial researchers will continue to use sorting techniques for some time to come. Although this paper provides an in-depth study of the effect of one particular technique, there are other techniques that could potentially be analyzed using the same methodology. Specifically, the result in this paper relies on the fact that the total sum of the squared variation of a sample is fixed. The sorting mechanism determines what part of this is attributable to the within-sample variation and what part is attributable to the between-sample variation. This apportioning of the variation directly affects the power and size of the statistical test, thereby affecting the statistical inferences. This basic idea can, in principal, be applied to other sorting techniques. For example, Lo and MacKinlay (1990) have argued against using variables identified within the sample as criteria for forming portfolios. Viewed from the perspective of this paper, what Lo and MacKinlay argue is that by using these variables to form the portfolios, the between-sample variance is increased, or to put this another way, the technique increases the variation in realized returns across portfolios. Because the sorting variables are identified within the sample, part of this variation might be spurious. Thus the inability of an asset pricing model to explain the between-sample variation might result from this data mining bias rather than reflecting a fundamental problem with the model.

The use of sorted data in empirical studies has become so widespread that this procedure elicits few econometric queries, and so little econometric justification for the technique is offered. This study emphasizes, at least in the context of one sorting procedure, why such justification might be important.

Appendix

Proof of Lemma 1: Take a partition with n fractiles. Then from the definition of the variance,

$$\sigma_{j(n)}^2 + E_{j(n)}^2 = \int_{i \in j(n)} (E[R_i])^2 \, dF^{j(n)}. \tag{A1}$$

Multiplying both sides by $Q_{j(n)}$ and then summing over $j(n)$ provides

$$\begin{aligned} \sum_{j(n)=1}^{n} Q_{j(n)}(\sigma_{j(n)}^2 + E_{j(n)}^2) &= \sum_{j(n)=1}^{n} \int_{i \in j(n)} (E[R_{it}])^2 Q_{j(n)} \, dF^{j(n)} \\ &= \sum_{j(n)=1}^{n} \int_{i \in j(n)} (E[R_{it}])^2 \, dF \\ &= \int_{i} (E[R_{it}])^2 \, dF \\ &= \sigma^2 + \mu^2. \end{aligned} \tag{A2}$$

Rearranging terms,

$$\begin{aligned} \sigma^2 - \sum_{j(n)=1}^{n} Q_{j(n)} \sigma_{j(n)}^2 &= \sum_{j(n)=1}^{n} Q_{j(n)}(E_{j(n)}^2 - \mu^2) \\ &= \sum_{j(n)=1}^{n} Q_{j(n)}((\mu + (E_{j(n)} - \mu))^2 - \mu^2) \\ &= \sum_{j(n)=1}^{n} Q_{j(n)}(E_{j(n)} - \mu)^2 + 2 \sum_{j(n)=1}^{n} Q_{j(n)}(\mu E_{j(n)} - \mu^2) \\ &= \sum_{j(n)=1}^{n} Q_{j(n)}(E_{j(n)} - \mu)^2 + 2\mu \sum_{j(n)=1}^{n} Q_{j(n)} E_{j(n)} - 2\mu^2 \\ &= \sum_{j(n)=1}^{n} Q_{j(n)}(E_{j(n)} - \mu)^2 = \sum_{j(n)=1}^{n} Q_{j(n)}(E_{j(n)}^*)^2, \end{aligned} \tag{A3}$$

where $E^*_{j(n)} \equiv E_{j(n)} - \mu$. This implies that

$$\sum_{j(n)=1}^{n} Q_{j(n)}\sigma^2_{j(n)} = \sigma^2 - \sum_{j(n)=1}^{n} Q_{j(n)}(E^*_{j(n)})^2 < \sigma^2, \tag{A4}$$

which completes the proof of the first part of the lemma.

For the second part, take another partition with m fractiles where $m > n$ and every fractile in the new partition is a subset of a fractile in the old partition: for every $j(m)$ there exists a $j(n)$ such that $j(m) \subseteq j(n)$. Now take the set of partitions that are a subset of $j(n)$:

$$\begin{aligned}
&\sum_{j(m)\subseteq j(n)} Q_{j(m)}(E^*_{j(m)})^2 \\
&\quad = \sum_{j(m)\subseteq j(n)} Q_{j(m)}(E^*_{j(n)} + (E^*_{j(m)} - E^*_{j(n)}))^2 \\
&\quad = \sum_{j(m)\subseteq j(n)} Q_{j(m)}(E^*_{j(m)} - E^*_{j(n)})^2 \\
&\qquad + \sum_{j(m)\subseteq j(n)} Q_{j(m)}((E^*_{j(n)})^2 + 2E^*_{j(n)}(E^*_{j(m)} - E^*_{j(n)})) \\
&\quad = \sum_{j(m)\subseteq j(n)} Q_{j(m)}(E^*_{j(m)} - E^*_{j(n)})^2 + 2E^*_{j(n)}\left(\sum_{j(m)\subseteq j(n)} Q_{j(m)}E^*_{j(m)}\right) \\
&\qquad - E^*_{j(n)} \sum_{j(m)\subseteq j(n)} Q_{j(m)}E^*_{j(n)} \\
&\quad = \sum_{j(m)\subseteq j(n)} Q_{j(m)}(E^*_{j(m)} - E^*_{j(n)})^2 + Q_{j(n)}(E^*_{j(n)})^2.
\end{aligned} \tag{A5}$$

Rearranging terms,

$$\left(\sum_{j(m)\subseteq j(n)} Q_{j(m)}(E^*_{j(m)})^2\right) - Q_{j(n)}(E^*_{j(n)})^2 = \sum_{j(m)\subseteq j(n)} Q_{j(m)}(E^*_{j(m)} - E^*_{j(n)})^2 \geq 0. \tag{A6}$$

Since $m > n$, this inequality is strict for at least one $j(n)$ so we have

$$\sum_{j(m)=1}^{m} Q_{j(m)}(E^*_{j(m)})^2 > \sum_{j(n)=1}^{n} Q_{j(n)}(E^*_{j(n)})^2. \tag{A7}$$

Since it is assumed that the $E^*_{j(n)}$ are distinct,

$$\lim_{n\to\infty} \sum_{j(n)=1}^{n} Q_{j(n)}(E^*_{j(n)})^2 = \sigma^2. \tag{A8}$$

These two facts and equation (A4) imply that for all $j(n)$

$$\lim_{n\to\infty} \sigma^2_{j(n)} = 0, \tag{A9}$$

which delivers the result. Q.E.D.

Proof of Proposition 1: Consider a grouping with n fractiles. By part 1 of Lemma 1,

$$\sigma^2 > \sum_{j(n)=1}^{n} Q_{j(n)} \sigma^2_{j(n)} \geq \min_{j(n)}\{\sigma_{j(n)}\} \sum_{j(n)=1}^{n} Q_{j(n)} = \min_{j(n)}\{\sigma_{j(n)}\}. \tag{A10}$$

Define

$$\Phi \equiv \sigma^2 - \min_{j(n)}\{\sigma_{j(n)}\} > 0, \tag{A11}$$

and take any grouping such that

$$\max_{j(n)}\{\sigma_{j(n)}\} - \min_{j(n)}\{\sigma_{j(n)}\} < \Phi. \tag{A12}$$

(At least one such grouping exists because the grouping in which $\sigma_{j(n)} = K$ for all j trivially satisfies this condition.) Then,

$$\sigma^2 = \min_{j(n)}\{\sigma_{j(n)}\} + \Phi > \max_{j(n)}\{\sigma_{j(n)}\}, \tag{A13}$$

which proves the proposition. Q.E.D.

Proof of Proposition 2: Consider the set of all possible series of groupings. Take the subset of series for which, for each series in this subset, there exists an m such that for all $n > m$,

$$\max_{j(n)}\{\sigma_{j(n)}\} - \min_{j(n)}\{\sigma_{j(n)}\} < \Delta, \tag{A14}$$

where $\Delta < \theta/\sqrt{1/\eta - 1}$. By part 2 of Lemma 1, for $\Gamma = \theta/\sqrt{1/\eta - 1} - \Delta$ there exists an N such that, for all $n > N$,

$$\Gamma > \sum_{j(n)=1}^{n} Q_{j(n)} \sigma^2_{j(n)} \geq \min_{j(n)}\{\sigma_{j(n)}\}. \tag{A15}$$

So,

$$\Gamma + \Delta > \min_{j(n)}\{\sigma_{j(n)}\} + \Delta > \max_{j(n)}\{\sigma_{j(n)}\}, \tag{A16}$$

which completes the proof. Q.E.D.

REFERENCES

Black, Fischer, Michael C. Jensen, and Myron Scholes, 1972, The capital asset pricing model: Some empirical tests; in Michael C. Jensen ed.: *Studies in the Theory of Capital Markets* (Praeger Publishers, New York).

Daniel, Kent, and Sheridan Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns, *Journal of Finance* 52, 1–34.

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–466.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Huang, Chi-fu, and Robert H. Litzenberger, 1988, *Foundations for Financial Economics* (North-Holland, New York).

Kim, Dongcheol, 1996, The errors in the variables problem in the cross-section of expected stock returns, *Journal of Finance* 50, 1605–1634.

Liang, Bing, 2000, Portfolio formation, measurement errors, and beta shifts: A random sampling approach, *Journal of Financial Research*, forthcoming.

Litzenberger, Robert H., and Krishna Ramaswamy, 1979, The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence, *Journal of Financial Economics* 7, 163–196.

Litzenberger, Robert H., and Krishna Ramaswamy, 1980, Dividends, short selling restrictions, tax-induced investor clienteles and market equilibrium, *Journal of Finance* 37, 469–482.

Lo, Andrew W., and A. Craig MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3, 431–468.