Journal of Financial Economics 49 (1998) 345–373

Alternative factor specifications, security characteristics, and the cross-section of expected stock returns[1]

Michael J. Brennan[a,b], Tarun Chordia[c], Avanidhar Subrahmanyam[d,*]

[a] Anderson Graduate School of Management, University of California-Los Angeles, Los Angeles, CA 90095, USA

[b] London Business School, Sussex Place, Regents Park, London, NWI 4SA, UK

[c] Owen Graduate School of Management, Vanderbilt University, Nashville, TN 37203, USA

[d] Anderson Graduate School of Management, University of California at Los Angeles, Los Angeles, CA 90095, USA

Received 2 December 1996; received in revised form 17 December 1997

Abstract

We examine the relation between stock returns, measures of risk, and several non-risk security characteristics, including the book-to-market ratio, firm size, the stock price, the dividend yield, and lagged returns. Our primary objective is to determine whether non-risk characteristics have marginal explanatory power relative to the arbitrage pricing theory benchmark, with factors determined using, in turn, the Connor and Korajczyk (CK; 1988) and the Fama and French (FF; 1993b) approaches. Fama–MacBeth-type regressions using risk adjusted returns provide evidence of return momentum, size, and book-to-market effects, together with a significant and negative relation between returns and trading volume, even after accounting for the CK factors. When the analysis is repeated using the FF factors, we find that the size and book-to-market effects are attenuated, while the momentum and trading volume effects persist. In addition, Nasdaq stocks show significant underperformance after adjusting for risk using either method. © 1998 Elsevier Science S.A. All rights reserved.

JEL classification: G12; G14

Keywords: Asset pricing; Anomalies; Risk factors

* Corresponding author. Tel.: 310/825-5355; fax: 310/206-5455; e-mail: subra@anderson.ucla.edu.

[1] We are especially grateful to Eugene Fama (a referee), an anonymous referee and Bill Schwert (the editor) for insightful and constructive suggestions. We also thank Wayne Ferson, Ken French,

PII S0304-405X(98)00028-2

1. Introduction

Early empirical research on the determinants of expected stock returns was concerned with detecting an association between average returns on beta-sorted portfolios and their betas, as predicted by the capital asset pricing model (see, e.g., Black, et al., 1972). Subsequently, Gibbons (1982) and Stambaugh (1982) introduced statistical tests of the null hypothesis that expected returns are determined *solely* by betas.[2] Following the development of the arbitrage pricing theory (APT), a similar series of tests was conducted, in which proxies for the APT factors and factor loadings replaced the market portfolio and betas of the CAPM.[3] Starting with the work of Black and Scholes (1974), Basu (1977), and Banz (1981), researchers began to test these asset pricing models against specific alternatives; these alternative hypotheses posited that expected returns on securities, instead of being determined *solely* by the risk characteristics of the securities, as measured by betas or factor loadings, were also affected by non-risk security characteristics such as size, book-to-market ratios, dividend yields, and earnings-price ratios. The role of some of these non-risk characteristics can be accounted for by frictions within the rational pricing paradigm, or could possibly be accounted for by their statistical properties as proxies for expected returns. However, the role of some other characteristics such as firm size has remained more elusive, so that their apparent importance for expected returns leaves the empirical validity of the rational asset pricing paradigm open to question.

In an important series of papers, Fama and French (FF) (1992a, b, 1993b, 1996) have provided evidence for the continuing validity of the rational pricing paradigm by showing that, with the exception of the momentum strategy of Jegadeesh and Titman (1993, 1995), the cross-sectional variation in expected returns associated with these non-risk characteristics can be captured by only

Footnote 1 continued
Will Goetzmann, Craig Holden, Ravi Jagannathan, Bob Jennings, Bruce Lehmann, Josef Lakonishok, Richard Roll, participants at the 1997 Meetings of the Western Finance Association, the 1997 UCLA/USC/UC Irvine conference, the November 1997 Asset Pricing Meeting of the National Bureau of Economic Research, the Atlanta Forum, and seminars at Columbia, Indiana, Florida, New York, Tulane, and Yale Universities; Eugene Fama and Ken French for providing part of the data used in this study; and Christoph Schenzler for excellent programming assistance. The second author acknowledges support from the Dean's Fund for Research and the Financial Markets Research Center at Vanderbilt University. We are responsible for remaining errors. This paper was formerly titled 'A Re-Examination of Security Return Anomalies'.

[2] Gibbons (1982), Stambaugh (1982).

[3] Roll and Ross (1980), Brown and Weinstein (1983), Shanken (1987), and Lehmann and Modest (1988).

two characteristics, namely the firm's size and its book-to-market ratio;[4] and that, moreover (FF, 1993b) these firm characteristics proxy for the security's loadings on priced factors. They show that the firm size and book-to-market effects can be accounted for within a three-factor model in which the factors are the returns on the market portfolio, and on two zero net-investment portfolios, one of which is long in high book-to-market and short in low book-to-market securities (HML), and the other of which is long in small firms and short in large firms (SMB).[5]

An important feature of much of this empirical research on asset pricing is that the analyzed returns are those on portfolios constructed by sorting securities on some criterion of interest. Portfolios are formed either to mitigate problems caused by using estimated betas as independent variables in a two-step estimation procedure or, when a one-step estimation procedure is used, to allow estimation of the covariance matrix of residual returns. This causes two quite different types of problem. First, as Roll (1977) has pointed out, the portfolio formation process, by concealing possibly return relevant security characteristics within portfolio averages, may make it difficult to reject the null hypothesis of no effect on security returns. Lo and MacKinlay (1990) make an almost precisely opposite point, that if the researcher forms portfolios on the basis of characteristics which prior empirical research has found to be related to average returns, he will be inclined to reject the null hypothesis too often due to a 'data-snooping' bias.[6] The resulting problem of inference is illustrated in FF (1996) and Brennan et al. (1996), who present results for six and seven sets of portfolios, respectively, and obtain quite different results depending on the criteria used in portfolio formation.

In this paper we investigate the extent to which expected returns can be explained by risk factors rather than by non-risk characteristics. Our approach differs from that of FF in three principal ways. First, rather than specifying the

[4] Fama and French (1992a) show that firm size and the ratio of book to market equity capture the cross-sectional relation between average returns and earnings yield and leverage.

[5] Daniel and Titman (1997) assert that portfolios of firms that have similar characteristics (size and book-to-market), but different loadings on the Fama French factors, have similar average returns, and use this finding to conclude that these security characteristics have an independent influence on expected returns.

[6] Table 5 of Lo and MacKinlay (1990) shows that if the R^2 between the sorting characteristic used to form portfolios and the estimated α's is 0.005, then the probability that a standard F-test will reject the null that the α's are jointly zero at the 5% level is 11.8% if 1000 securities are sorted into 10 portfolios of 100 securities, even though the underlying data satisfy the null hypothesis. If the R^2 is 0.01 the size of a 5% test rises to 36.7% for 1000 securities sorted into 10 portfolios of 100 securities, even though the underlying data satisfy the null hypothesis. If no portfolio aggregation had been performed the size of these tests would be 5%!

risk factors a priori, we follow the intuition of the APT, that the risk factors should be those which capture the variation of returns in large well-diversified portfolios, and use the principal components approach of Connor and Korajczyk (1988) (henceforth CK) to estimate risk factors. We then repeat the analysis using the FF (1993b) factors.[7] Thus, our null hypothesis is that expected returns are determined by the APT with risk factors obtained using the Connor and Korajczyk or the Fama and French approach. Secondly, rather than limiting ourselves to the set of firm characteristics that Fama and French have found to be associated with average returns, notably size and book-to-market ratio, we estimate simultaneously the marginal effects of eight firm characteristics, including dividend yield, and measures of market liquidity such as share price and trading volume, as well as lagged returns. We are able to consider these several characteristics simultaneously because, thirdly, instead of examining the returns on portfolios, we examine the *risk-adjusted* returns on individual securities.[8] Under the null hypothesis, these risk-adjusted returns should be independent of other (non-risk) security characteristics. Not only does this approach allow us to consider the effects of a large number of firm characteristics simultaneously, but it also avoids the data-snooping biases that are inherent in the portfolio-based approaches as discussed above.[9] Our approach also avoids the errors-in-variables bias created by errors in estimating factor loadings, since errors in the factor loadings are impounded in the dependent variable. The costs of this approach are that it imposes the assumption that the zero-beta return equals the risk-free rate, and it incorporates the prediction of the APT that the realized reward per unit of loading on a given factor is equal to the realized return on the underlying factor portfolio.

When we use only size, book-to-market, and lagged returns as the explanatory variables, we find that these variables are significantly related to expected returns even after risk-adjustment using the CK factors. When the analysis is repeated using the FF portfolios as factors, the size and book-to-market effects

[7] Campbell (1996), using the intuition of Merton's (1973) intertemporal CAPM, argues that 'priced factors should be found not by running a factor analysis on the covariance matrix of returns… Instead, innovations in variables that have been found to forecast stock returns and labor income should be used'. It seems likely to us that variables that have a significant effect on the future investment opportunity set are also likely to have a significant effect on contemporaneous returns, so that their traces will be evident in the covariance matrix of returns.

[8] Papers that use risk- *unadjusted* returns for cross-sectional analyses on individual securities include FF (1992a), Litzenberger and Ramaswamy (1979), Miller and Scholes (1982), and Lehmann (1990).

[9] Of course, we are guilty of data-snooping in a different sense: The security characteristics we have chosen to consider are motivated by previous results. But we do avoid the aggravation of the problem caused by sorting to form portfolios. Ferson et al. (1998) also point out the pitfalls in using attribute sorted portfolios as risk factors.

are attenuated by a factor of about 1/3, and their significance is weakened as well. Expanding the set of explanatory variables, we find that a return-momentum effect persists, and also that there is a negative and significant relation between returns and trading volume, regardless of whether the risk-adjustment is done with the CK factors or the FF factors. In addition, the introduction of trading volume makes the coefficient of the firm size variable positive and significant. The dividend yield variable is significant with the CK factors but not with the FF factors.

The fact that the 'non-risk' firm characteristics are significant explanators of the 'risk-adjusted' returns implies either that the risk adjustment is incomplete, or that returns are affected by other factors than risk. While the dividend yield effect is present only under the CK risk-adjustment procedure, the trading volume effect we find is rather robust, in that it is present for both types of risk-adjustment, as well as in risk-unadjusted returns; this supports the notion that this variable is acting as a proxy for the liquidity of the market in the firm's shares,[10] rather than as a proxy for the loading on some priced risk factor that is not included in the analysis.

In order to account for the fact that trading volume is measured differently on NYSE/AMEX and Nasdaq, we include separate variables for Nasdaq and NYSE volume. Since the Nasdaq volume is not significant and Reinganum (1990) and Loughran (1993) provide evidence of a 'Nasdaq' effect, we include a dummy variable for Nasdaq membership. We then find that dollar volume is strongly negatively associated with returns for both exchanges, but find that holding constant their factor loadings and other characteristics Nasdaq stocks *underperform* by about 10% per year.

We find that the five CK factors offer a risk-return trade-off that is comparable to that offered by the three FF factors in the sense that the overall squared Sharpe ratios are close; for both sets of factors the null hypothesis that the reward-for-risk ratio equals zero can be rejected at better than the 1% level of significance. However, our analysis suggests that the two sets of factors are not equivalent. Indeed, we find using Gibbons et al. (1989) intercept tests that neither set of factors price the other, though there is evidence that CK factors are priced better by the FF factors than are the FF factors by the CK factors.

The remainder of the paper is organized as follows. In Section 2 we describe the empirical hypotheses we test. In Section 3 the data are described and in Section 4 the statistical model is presented. In Section 5, we present the regression results, while in Section 6, we compare the FF and CK factors, and Section 7 concludes.

---

[10] Glosten and Harris (1988) and Brennan and Subrahmanyam (1995) show that trading volume is a major determinant of market liquidity.

2. Hypotheses

Our null hypothesis is an L-factor version of the APT which implies that the expected excess return on security j is determined solely by the loadings of the security's return on the L factors, $\beta_{jk}(k = 1, \ldots, L)$. Consider the following equation:

$$\mathrm{E}(\tilde{R}_j) - R_F = c_0 + \sum_{k=1}^{L} \lambda_k \beta_{jk} + \sum_{m=1}^{M} c_m Z_{mj}, \tag{1}$$

where $\tilde{R}_j$ is the return on security j, R_F is the risk free interest rate, β_{jk} is the loading of security j on factor k, λ_k is the risk premium associated with factor k, Z_{mj} $(m = 1, \ldots, M)$ is the value of (non-risk) characteristic m for security j, and c_m is the premium per unit of characteristic m. Our null hypothesis is that $c_m = 0$ $(m = 0, 1, \ldots, M)$. We include eight non-risk security characteristics (including three momentum-based lagged return variables) as possible determinants of expected returns.

The risk factors are initially taken to be the first five (asymptotic) principal components of excess stock returns estimated over the sample period, and, in turn, the three FF factors. In deciding which non-risk firm characteristics to include as possible determinants of expected returns, attention was given to those variables that had been found to be important in prior studies, as well as those for which there exists a theoretical rationale. Thus firm size is included because of the importance of assessing whether the 'small firm effect' (see Banz, 1981; FF, 1992a) persists after accounting for the five risk factors[11] and other firm characteristics. We also include the ratio of book-to-market equity because this has been found to be strongly associated with average returns (see FF, 1992a; Lakonishok et al., 1994). It has been hypothesized that the low price effect documented by Miller and Scholes (1982) reflects the fact that firms with low prices are often in financial distress, and that financial institutions may be reluctant to invest in them on account of the prudent man rule.[12] Therefore we include the reciprocal of share price as a possible determinant of expected returns.

Amihud and Mendelson (1986) and Brennan and Subrahmanyam (1996) suggest that expected returns are affected by liquidity. Amihud and Mendelson use the bid–ask spread as a measure of liquidity. However, the spread is available only annually, and only for NYSE/AMEX stocks. Brennan and

[11] Lehmann and Modest (1988) found that their implementation of a five-factor APT was unable to account for the size anomaly.

[12] Falkenstein (1996) shows that mutual funds 'show an aversion to low-price stocks'.

Subrahmanyam, on the other hand, use the fixed and variable components of trading costs as measures of liquidity. Since their measures require intraday data, which is available only after 1983, their sample period is short. In our study, we include the dollar volume of trading because this variable is associated with liquidity,[13] and because Petersen and Fialkowski (1994) find that the quoted spread is only loosely associated with the effective spread; therefore it is possible that trading volume provides a better measure of liquidity than the bid-ask spread. Further, dollar volume is available monthly, and thus may allow a more powerful test of the liquidity hypothesis.

We include dividend yield because Brennan (1970) suggests that differential taxation of dividends and capital gains could make this variable relevant, and the resulting empirical work of Litzenberger and Ramaswamy (1979) and Miller and Scholes (1978, 1982) has been inconclusive. Finally, we include lagged return variables because Jegadeesh and Titman (1993) have shown these to be relevant, and by including them we should improve the efficiency of the estimates of the coefficients of the other variables.

3. Data

The basic data consist of monthly returns and other characteristics for a sample of the common stock of companies for the period January 1966 to December 1995.[14] To be included in the sample for a given month a stock had to satisfy the following criteria: (1) Its return in the current month and in 24 of the previous 60 months be available from CRSP, and sufficient data be available to calculate the size, price, dollar volume, and dividend yield as of the previous month; (2) Sufficient data be available on the COMPUSTAT tapes to calculate the book-to-market ratio as of December of the previous year. As per Fama and French (1992) we excluded financial firms from our sample. This screening process yielded an average of 2457 stocks per month.

For each stock the following variables were calculated each month as follows:

SIZE – the natural logarithm of the market value of the equity of the firm as of the end of the second to last month.

BM – the natural logarithm of the ratio of the book value of equity plus deferred taxes to the market value of equity, using the end of the previous year

[13] Several studies (e.g., Stoll (1978)) find trading volume to be the most important determinant of the bid-ask spread, and Brennan and Subrahmanyam (1995) find that it is a major determinant of their measure of liquidity.

[14] The observation period began in January 1966 because the FF factors are available only from July 1963 onwards, and we required enough lag time to allow loadings to be estimated reliably from past factor realizations.

market and book values. As in FF (1992a), the value of BM for July of year t to June of year $t+1$ was computed using accounting data at the end of year $t-1$, and book-to-market ratio values greater than the 0.995 fractile or less than the 0.005 fractile were set equal to the 0.995 and 0.005 fractile values, respectively.

DVOL – the natural logarithm of the dollar volume of trading in the security in the second to last month.

PRICE – the natural logarithm of the reciprocal of the share price as reported at the end of the second to last month.

YLD – the dividend yield as measured by the sum of all dividends paid over the previous 12 months, divided by the share price at the end of the second to last month.

RET2–3 – the natural logarithm of the cumulative return over the two months ending at the beginning of the previous month.

RET4–6 – the natural logarithm of the cumulative return over the three months ending three months previously.

RET7–12 – the natural logarithm of the cumulative return over the 6 months ending 6 months previously.

The lagged return variables were constructed to exclude the return during the immediate prior month in order to avoid any spurious association between the prior month return and the current month return caused by thin trading or bid–ask spread effects. In addition, all variables involving the price level were lagged by one additional month in order to preclude the possibility that a linear combination of the lagged return variables, the book-to-market variable (which is related to the price level in the previous year), and the reciprocal of the price level could provide a noisy estimate of the return in the previous month, thus leading to biases because of bid-ask effects and thin trading.[15]

Table 1 reports the time-series averages of the cross-sectional means, medians, and standard deviations of the raw (i.e., unlogged) security characteristics, and displays the summary statistics associated with both trimmed and untrimmed values of the book-to-market ratio.

The variables display considerable skewness. Therefore, in our empirical analysis we employ logarithmic transforms of all these variables except the dividend yield (which may be zero). Finally, for all of the regressions reported below, the transformed firm characteristics variables for a given month were expressed as deviations from their cross-sectional means for that month; this implies that the average security will have values of each non-risk characteristic that are equal to zero, so that under both the null and the alternative hypotheses its expected return will be determined solely by its risk characteristics. Table 2

[15] See Jegadeesh (1990). It is easy to show that thin trading will cause risk-adjusted returns to exhibit first-order negative serial correlation.

Table 1
Summary statistics

The summary statistics represent the time-series averages of cross-sectional means for an average of 2457 stocks over 360 months from Jan. 1966 through Dec. 1995. Each stock satisfies the following criteria: (1) Its return in the current month and in 24 of the previous 60 months be available from CRSP, and sufficient data be available to calculate the size, price, dollar volume, and dividend yield as of the previous month; and (2) Sufficient data be available on the COMPUSTAT tapes to calculate the book to market ratio as of December of the previous year. The row titled book-to-market ratio (trimmed) provides summary statistics for the book-to-market ratio after values greater than the 0.995 fractile or less than the 0.005 fractile are set to equal the 0.995 and 0.005 fractile values, respectively.

Variable	Mean	Median	Std. Dev.
Firm size ($ billion)	0.436	0.061	0.099
Book-to-market ratio	2.074	0.836	0.715
Book-to-market ratio (trimmed)	1.250	0.836	0.534
Dollar-trading-volume ($ million per month)	17.627	1.925	13.014
Share price ($)	19.804	15.039	6.767
Dividend yield (%)	2.51	1.67	0.950

reports the averages of the month by month cross-sectional correlations of the transformed variables that we use in our analysis. The largest correlations are between SIZE and DVOL and SIZE and PRICE. The other correlations are smaller than 0.40 in absolute value.

The five CK factors were estimated by the asymptotic principal components technique developed by Connor and Korajczyk (1988)[16] applied to returns in excess of the risk-free rate on all securities listed continuously over the estimation period, where the risk-free rate was taken as the 1 month risk free rate from the CRSP bond files. In order to keep the estimation process computationally manageable, the factors were estimated separately over each of two over-lapping subperiods: July 1963 to December 1979 and January 1975 to December 1995. The three FF factors are the market portfolio, SMB which is intended to mimick the performance of a portfolio that is long in small firms and short in large firms, and HML which is intended to mimic the performance of a portfolio which is

[16] Connor and Korajczyk (1993) 'find evidence for one to six pervasive factors generating returns on the NYSE and AMEX over the period 1967 to 1991'.

Table 2
Correlation matrix of transformed firm characteristics

This table presents time-series of monthly cross-sectional correlations between the transformed firm characteristics used in pricing regressions. The variable relate to an average of 2457 stock over 360 months from Jan 1966 through Dec. 1995. RETURN denotes the excess monthly return, i.e., the raw return less the risk-free return. SIZE represents to logarithm of the market capitalization on of firms in billions of dollars. BM is the logarithm of the ratio of book value of equity plus deferred taxes to market capitalization, with the expection that book-to-market ratio values greater than the 0.995 fracticle or less than the 0.005 fractile are set to equal the 0.995 and 0.005 fracticle values, respectively. DVOL is the logarithm of the dollar trading volume. PRICE is the logarithm of the reciprocal of the share price. YLD is the logarithm of the dividend yield; RET2–3, RET4–6, RET7–12 equal the logarithms of the cumulative returns over the second through third, fourth through sixth, and seventh through 12th months prior to the current month, respectively.

	RETURN	SIZE	BM	DVOL	PRICE	YLD	RET2–3	RET4–6	RET7–12
RETURN	1.00	– 0.010	0.030	– 0.006	0.004	0.007	0.009	0.018	0.024
SIZE	– 0.010	1.00	– 0.238	0.753	– 0.790	0.084	0.011	0.013	0.017
BM	0.030	– 0.238	1.00	– 0.146	0.158	0.144	– 0.015	0.040	0.047
DVOL	– 0.006	0.753	– 0.146	1.00	– 0.387	0.028	0.048	0.049	0.067
PRICE	0.004	– 0.790	0.156	– 0.387	1.00	– 0.196	– 0.188	– 0.127	– 0.145
YLD	0.007	0.084	0.144	0.028	– 0.196	1.00	– 0.044	– 0.043	– 0.042
RET2–3	0.009	0.011	– 0.015	0.048	– 0.188	– 0.044	1.00	– 0.005	0.030
RET4–6	0.018	0.013	0.040	0.049	– 0.127	– 0.043	– 0.005	1.00	0.038
RET7–12	0.024	0.017	0.047	0.067	– 0.145	– 0.042	0.030	0.038	1.00

long high book-to-market equity firms and short low book-to-market equity firms.[17]

4. Statistical model

As we have argued above, empirical findings based on the returns on *portfolios* are hard to interpret. Therefore, we report the results from analyzing the returns on individual securities. The null hypothesis against which we evaluate the influence of the non-risk security characteristics is an L-factor APT.

[17] As noted in Footnote 14, the FF factors are available only from July 1963 onwards. This is why we start the estimation period for the CK factors in July 1963 as well.

Thus, assume that returns are generated by an L-factor approximate factor model:[18]

$$\tilde{R}_{jt} = \mathrm{E}(\tilde{R}_{jt}) + \sum_{k=1}^{L} \beta_{jk}\tilde{f}_{kt} + \tilde{e}_{jt}. \tag{2}$$

Then the exact or equilibrium version of the APT, in which the market portfolio is well-diversified with respect to the factors (Connor, 1984; Shanken, 1985, 1987), implies that expected returns may be written as

$$\mathrm{E}[\tilde{R}_{jt}] - R_{Ft} = \sum_{k=1}^{L} \lambda_{kt}\beta_{jk}, \tag{3}$$

where R_{Ft} is the return on the riskless asset, and λ_{kt} is the risk premium for factor k. Substituting from Eq. (3) in Eq. (2), the APT implies that realized returns are given by

$$\tilde{R}_{jt} - R_{Ft} = \sum_{k=1}^{L} \beta_{jk}\tilde{F}_{kt} + \tilde{e}_{jt}, \tag{4}$$

where $\tilde{F}_{kt} \equiv \lambda_{kt} + \tilde{f}_{kt}$ is the sum of the factor realization and its associated risk premium. Our goal is to test whether security characteristics have incremental explanatory power for returns relative to the five-factor CK benchmark or the three-factor FF benchmark.

A standard application of the Fama–MacBeth (1973) procedure would involve estimation of the following equation:

$$\tilde{R}_{jt} - R_{Ft} = c_0 + \sum_{k=1}^{L} \beta_{jk}\tilde{f}_{kt} + \sum_{m=1}^{M} c_m Z_{mjt} + \tilde{e}_{jt}, \tag{5}$$

where Z_{mjt} is the value of characteristic m for security j in month t. Under the null hypothesis that expected returns depend only on the risk characteristics of the returns, as represented by β_{jk}, the loadings on the CK or FF factors, the coefficients c_m ($m = 1, \ldots, M$) will be equal to zero. This hypothesis can be tested in principle by estimating the factor loadings for each month using prior data, estimating a cross-section regression for each month in which the independent variables are the factor loadings and non-risk characteristics, and then averaging the monthly coefficient estimates over time and calculating their time-series standard errors. This standard Fama–MacBeth approach, however, presents problems because the factor loadings are measured with error. One method of dealing with this measurement error problem is to use the information from the first-stage regressions (in which the factor loadings are estimated)

[18] See Connor and Korajczyk (1988) for the definition of an approximate factor model.

to correct the coefficient estimates in the second stage regressions.[19] Our approach to correct the bias, however, does not rely on information taken from the first stage regressions.

First, each year, from 1966 to 1995, factor loadings, β_{jk}, were estimated for all securities that had at least 24 return observations over the prior 60 months, with the qualification that since our factor estimation begins in July 1963, the factor loadings in the first month of the regression period (January 1966) were estimated from 30 observations per factor, the next month, 31, and so on till the 60 month level was reached from which point the observation interval was kept constant at 60 months.[20] In order to allow for thin trading, we used the Dimson (1979) procedure with one lag to adjust the estimated factor loadings. The estimated risk-adjusted return on each of the securities, $\tilde{R}^*_{jt}$, for each month t of the following year was then calculated as:

$$\tilde{R}^*_{jt} \equiv \tilde{R}_{jt} - R_{Ft} - \sum_{k=1}^{L} \hat{\beta}_{jk} \tilde{F}_{kt}. \tag{6}$$

As pointed out in the introduction, our risk adjustment procedure imposes the assumptions that the zero-beta equals the risk-free rate, and that the APT factor premium is equal to the excess return on the factor. The risk-adjusted returns from Eq. (6) constitute the raw material for the estimates that we present below of the equation:

$$\tilde{R}^*_{jt} = c_0 + \sum_{m=1}^{M} c_m Z_{mjt} + \tilde{e}'_{jt} \tag{7}$$

Note that the error term in Eq. (7) is different from that in Eq. (4), because the error in Eq. (7) also contains terms arising from the measurement error associated with the factor loadings. We show how this measurement error affects our estimation in the discussion that follows.

We first calculate an estimate of the vector of characteristic rewards $\hat{\boldsymbol{c}}_t$ each month from a simple OLS regression:

$$\hat{\boldsymbol{c}}_t = (\boldsymbol{Z}'_t \boldsymbol{Z}_t)^{-1} \boldsymbol{Z}'_t \boldsymbol{R}^*_t, \tag{8}$$

where $\boldsymbol{Z}_t$ is the vector of firm characteristics in month t and $\boldsymbol{R}^*_t$ is the vector of estimated risk-adjusted returns. Note that although the factor loadings, β_{jk}, are estimated with error, this error affects only the dependent variable, $\boldsymbol{R}^*_t$, and

[19] This is the approach followed by Litzenberger and Ramaswamy (1979) and Lehmann (1990).

[20] We have one set of factors for each of the two overlapping subperiods; since there is no correspondence between factor k in the two subperiods, care was taken to ensure that the factors used for risk-adjustment were the same as those for which the factor loadings were estimated.

while the factor loadings will be correlated with the security characteristics, $\boldsymbol{Z}_t$, there is no a priori reason to believe that errors in the estimated loadings will be correlated with the security characteristics, so the estimated coefficient vector, $\hat{\boldsymbol{c}}_t$, is unbiased under the null hypothesis.

For each characteristic, m ($m = 0, 1, \ldots, M$) (including the constant term) the coefficient estimates, for each month from January 1966 to December 1995, are aggregated into an overall estimate in one of two ways. The first, which we call the raw estimate, is given by

$$\hat{c}_{mr} = (j'j)^{-1} j' \hat{\boldsymbol{c}}_m, \tag{9}$$

where j is the unit vector and $\hat{\boldsymbol{c}}_m$ is the vector of monthly estimates of c_m. Thus, Eq. (9) represents the time-series average of the coefficients associated with the characteristics: it is simply the standard Fama–MacBeth estimator except that the dependent variable is the risk-adjusted return, calculated using either the CK or the FF approach. While there is no a priori reason to believe that the errors in the estimated factor loadings will be correlated with the security characteristics, $\boldsymbol{Z}_t$, to the extent that they are correlated, the monthly estimates of the coefficients of the firm characteristics, $\hat{c}_{mt}$, will be correlated with the factor realizations, and therefore the mean of these estimates which is the Fama–Macbeth estimator will be biased by an amount that depends on the mean factor realizations. Therefore, as a check on the robustness of our results, a purged estimator, $\hat{c}_{mp}$, was obtained for each of the characteristics as the constant term from the regression of the monthly coefficient estimates on the time series of CK or FF factor realizations. This estimator, which was first developed by Black and Scholes (1974), purges the monthly estimates of the factor dependent component, is given by

$$\hat{c}_{mp} = \boldsymbol{e}'(\boldsymbol{F}^{*\prime}\boldsymbol{F}^{*})^{-1}\boldsymbol{F}^{*\prime}\hat{\boldsymbol{c}}_m, \tag{10}$$

where $\boldsymbol{e}$ is a 6-element vector $[1\ 0\ 0\ 0\ 0\ 0]'$ which serves to pick out the constant of the regression, and $\boldsymbol{F}^*$ is the matrix of factor portfolio returns augmented by a vector of ones. To see that the purged estimator is unbiased even when the errors in the factor loading estimates are correlated with the characteristics, $\boldsymbol{Z}$, denote the risk-adjusted return under the true factor loadings as $\tilde{R}^T_{jt}$. Then, from Eq. (6), we have

$$\tilde{R}^*_{jt} = \tilde{R}^T_{jt} + \sum_{k=1}^{L} \hat{u}_{jk} F_{kt},$$

where $u_{jk} \equiv \beta_{jk} - \hat{\beta}_{jk}$ is the measurement error in the kth factor loading for security j. Letting $\boldsymbol{c}$ and u be the true coefficient vector of the characteristics and the measurement error matrix, respectively, and $\boldsymbol{F}_t$ be the vector of factor observations in month t, the regression of risk-adjusted returns in month t on the security characteristics yields the following coefficient vector:

$$\hat{\boldsymbol{c}}_t = \boldsymbol{c} + \boldsymbol{F}_t[(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{u}]_t.$$

Thus, the intercept from the regression of $\hat{\boldsymbol{c}}_t$ on $\boldsymbol{F}_t$ will be an unbiased estimate of $\boldsymbol{c}$ so long as the factor realizations are serially uncorrelated.

In sum, $\hat{c}_{mr}$, represents the standard Fama–MacBeth estimator, and $\hat{c}_{mp}$ represents the constant from the OLS regression of the month-by-month Fama–MacBeth estimates on the factor portfolio returns for the purged estimator.[21] The standard error of the estimate is taken from the time series of monthly estimates in the case of the raw estimator, $\hat{c}_{mr}$, and from the standard error of the intercept from the OLS regression in the case of the purged estimator, $\hat{c}_{mp}$. As Shanken (1992) points out, the standard errors of the coefficients yielded by the standard Fama–MacBeth approach are understated because they ignore the additional variation induced by the estimation error in the factor loadings. We show in Section 6, however, that the magnitude of this understatement is small for our sample, and does not affect our basic conclusions.

5. Regression analysis

5.1. Results

To begin our analysis we present the results of Fama–MacBeth regressions of excess (risk-*un*adjusted) returns on characteristics which are best known to be associated with expected returns, namely, SIZE, BM, and the three lagged return variables. The results are reported in the first column of Table 3.

As can be seen, the coefficients of SIZE and BM are respectively negative and positive, and both are statistically significant, which is consistent with earlier studies such as FF (1992a). In addition, the coefficients of all of the three lagged return variables are positive, and two are strongly significant.

We now consider whether the relation between excess returns and SIZE, BM, and the lagged return variables is maintained when the returns are risk-adjusted returns using the two sets of factors. The raw and purged estimates of the characteristic rewards, $\hat{c}_{mr}$ and $\hat{c}_{mp}$, for risk-adjusted returns using the CK factors are reported in the second and third columns of Table 3. The coefficients of SIZE and BM are essentially unchanged by the risk-adjustment and are highly significant, and the coefficients of all of the three lagged return variables are positive and two of them are significant. There is little difference between the raw and purged estimates as we should expect if the factor loading errors are uncorrelated with the non-risk characteristics. For comparison, the results from

[21] Separate estimates are calculated corresponding to the two subperiods for which the principal components were estimated; these were then aggregated using precision weights.

Table 3
Fama–MacBeth regression estimates of Eq. (7) using individual security data

Coefficient estimates are time-series averages of cross-sectional OLS regressions. The dependent variable in the first column is simply the excess return, while in the second and third columns it is the excess returns risk-adjusted using the CK factors, and in the fourth and fifth columns it is the excess returns risk-adjusted using the FF factors (Dimson beats with one lag are used in each case). The independent variables are defined as follows; SIZE represents logarithm of the market capitalization of firms in billions of dollars. BM is the logarithm of the ratio of book value of equity plus deferred taxes to market capitalization, with the exception that book-to-market ratio values greater than the 0.995 fractile or less than the 0.005 fractile are set to equal the 0.995 and 0.005 fractile values, respectively, RET2–3, RET4–6, RET7–12 equal the logarithms of the cumulative returns over the second through third, fourth through sixth, and seventh through 12th months prior to the current month, respectively. The variables are measured as the deviation from the cross-sectional mean in each period. The estimates in the column labeled 'Raw' are the coefficients estimated using Eqs. (8) and (9), while those in the column labeled 'Purged' are from Eqs. (8) and (10). All coefficients are multiplied by 100. *t*-statistics are in parentheses.

	Excess returns	Risk-adjusted returns using the CK factors		Risk-adjusted returns using the FF factors	
		Raw	Purged	Raw	Purged
Intercept	0.735	0.412	0.101	0.099	0.642
	(2.36)	(0.63)	(1.85)	(1.45)	(0.96)
SIZE	− 0.140	− 0.157	− 0.150	− 0.106	− 0.096
	(2.70)	(4.81)	(4.60)	(2.95)	(2.63)
BM	0.295	0.271	0.264	0.173	0.171
	(4.52)	(4.95)	(4.85)	(3.44)	(3.41)
RET2–3	0.285	0.813	0.510	0.605	0.873
	(0.89)	(3.08)	(2.18)	(1.97)	(2.86)
RET4–6	0.624	0.847	0.693	0.881	1.145
	(2.19)	(3.23)	(3.21)	(3.24)	(4.31)
RET7–12	0.842	0.227	0.302	0.642	0.974
	(5.13)	(1.18)	(1.73)	(3.05)	(5.02)

risk-adjustment using the FF factors are reported in the last two columns. Both the size and book-to-market effects are now reduced by about one third, and their significance is attenuated as well. The lagged returns are highly significant, confirming FF (1996). Although for both sets of factors the intercept term is insignificantly different from zero as predicted by the null hypothesis, it is apparent that neither factor model provides a complete description of equilibrium returns.

In Table 4 we present the results of regressions that use the full set of characteristics: SIZE, BM, PRICE, DVOL, YLD, PRICE, as well as the lagged

Table 4
Fama–MacBeth regression estimates of Eq. (7) using individual security data

Coefficient estimates are time-series averages of cross-sectional OLS regressions. The dependent variable in the first column is simply the excess return, while in the second and third columns it is the excess returns risk-adjusted using the CK factors, and in the fourth and fifth columns it is the excess returns risk-adjusted using the FF factors (Dimson beats with one lag are used in each case). The independent variables are defined as follows; SIZE represents logarithm of the market capitalization of firms in billions of dollars. BM is the logarithm of the ratio of book value of equity plus deferred taxes to market capitalization, with the exception that book-to-market ratio values greater than the 0.995 fractile or less than the 0.005 fractile are set to equal the 0.995 and 0.005 fractile values, respectively. DVOL is the logarithm of the dollar trading volume. PRICE is the logarithm of the reciprocal of the share price. YLD is the logarithm, of the dividend yield; RET2–3, RET4–6, RET7–12 equal the logarithms of the cumulative returns over the second through third, fourth through sixth, and seventh through 12th months prior to the current month, respectively. NYDVOL is the value of DVOL if the stock trades on NYSE/AMEX, and zero otherwise; NADVOL is the value of DVOL if the stock trades on NASdaq; and zero otherwise. The estimates in the column labeled 'Raw' are the coefficients estimated using Eqs. (8) and (9), while those in the column labeled 'Purged' are from Eqs. (8) and (10). All coefficients are multiplied by 100. *t*-statistics are in parentheses. The variables are measured as the deviation from the cross-sectional mean in each period. The estimates in the column labeled 'Raw' are the coefficients estimated using Eqs. (8) and (9), while those in the column labeled 'Purged' are from Eqs. (8) and (9). All coefficients are multiplied by 100. *t*-statistics are in parentheses.

	Excess returns	Risk-adjusted returns using the Connor–Korajczyk factors		Risk-adjusted returns using the Fama–French factors	
		Raw	Purged	Raw	Purged
Intercept	0.707	0.004	0.092	0.071	0.035
	(2.25)	(0.06)	(1.69)	(1.02)	(0.51)
SIZE	0.092	0.116	0.143	0.122	0.106
	(1.56)	(2.57)	(3.15)	(2.84)	(2.46)
BM	0.246	0.201	0.188	0.128	0.129
	(5.02)	(4.12)	(3.85)	(2.87)	(2.90)
PRICE	0.196	0.166	0.153	0.109	0.013
	(1.87)	(1.97)	(1.78)	(1.16)	(0.14)
NYDVOL	– 0.130	– 0.190	– 0.199	– 0.162	– 0.173
	(2.68)	(5.02)	(5.34)	(4.17)	(4.38)
NADVOL	– 0.088	– 0.175	– 0.186	– 0.086	– 0.173
	(1.23)	(2.59)	(2.86)	(1.87)	(1.39)
YLD	0.215	1.778	2.371	0.643	0.327
	(0.13)	(1.82)	(3.33)	(0.57)	(0.28)
RET2–3	0.654	1.158	0.706	0.888	1.072
	(2.30)	(4.14)	(2.89)	(2.98)	(3.57)
RET4–6	0.789	1.062	0.849	1.006	1.205
	(3.26)	(3.99)	(3.85)	(2.81)	(4.59)
RET7–12	0.869	0.325	0.365	0.666	0.974
	(5.99)	(1.69)	(2.11)	(3.21)	(5.01)

return variables. Since trading volume is measured differently between NYSE/AMEX and Nasdaq,[22] we split DVOL into two variables: NYDVOL, which equals DVOL if the stock trades on NYSE/AMEX and zero otherwise, and NADVOL, which equals DVOL if the stock trades on Nasdaq and zero otherwise. The results using risk-unadjusted returns are presented in the first column of Table 4.

Now the coefficient of SIZE, which was previously negative and significant, is positive and no longer significant, whereas the coefficients of BM, NYDVOL, and all three lagged return variables are strongly significant.[23] These variables remain significant following risk-adjustment by the CK factors; the coefficient on SIZE and NADVOL now become significant. Particularly striking is the behavior of the coefficient on YLD which becomes large and positive after risk-adjustment. When risk-adjustment is carried out using the FF factors, YLD is insignificant though the coefficient on NYDVOL remains negative and significant. The BM effect is reduced by about 50%, although SIZE remains positive and significant. In summary, risk-adjustment by either set of factors leaves significant SIZE (positive), BM, and NYDVOL effects, as well as lagged return effects; the CK factors also leave a YLD effect and a NADVOL effect.

It is worth noting that the magnitudes of the coefficients on some of the Z variables increase substantially after risk adjustment by the CK factors – for example, the slopes on the volume variables, RET2-3, RET4-6, and particularly the one on YLD which increases by a factor of about nine. While the magnitudes of some of the coefficients also increase after risk-adjustment by the FF factors, the increase is less dramatic and the FF factors significantly reduce the magnitudes of the SIZE and BM coefficients.

The lack of significance of NADVOL, in contrast to the high level of significance of NYDVOL, in the FF regressions leaves the role of trading volume unclear. However, Reinganum (1990) finds that the average returns on NYSE securities exceed those of similar size firms listed on Nasdaq by about 6%

[22] It is well-known that Nasdaq volume is considered overstated relative to NYSE/AMEX volume, owing to the inclusion of inter-dealer trading on Nasdaq, and the requirement that most trades on Nasdaq must be submitted to a dealer, whereas crossing between brokers is not included in the reported trading volume on the other exchanges.

[23] We also performed a test of the null hypothesis that the coefficients of the characteristics in these regressions are *jointly* equal to zero. To do this, we calculated the Hotelling T^2 statistic, which, given N time-series observations of p coefficients, is defined as

$$T^2 = N[\bar{\gamma}' S^{-1} \bar{\gamma}],$$

where $\bar{\gamma}$ is the (time-series) mean vector of the coefficients, and S is the estimated variance covariance matrix of the coefficients. Under the null hypothesis, the T^2 statistic is distributed $[(N-1)_p/N-p]\, F_{p,\,N-p}$. We do not report the results of this test here, because in every regression that we performed, the null hypothesis that the coefficients jointly equal zero could be easily rejected, with p-values ranging from 0.02 to 10^{-20}.

Table 5
Fama–MacBeth regression estimates of Eq. (7) using individual security data, including dummy variable for Nasdaq stocks

Coefficient estimates are time-series averages of cross-sectional OLS regressions. The dependent variable in the first column is simply the excess return, while in the second and third columns it is the excess returns risk-adjusted using the CK factors, and in the fourth and fifth columns it is the excess returns risk-adjusted using the FF factors (Dimson beats with one lag are used in each case). The independent variables are defined as follows; SIZE represents logarithm of the market capitalization of firms in billions of dollars. BM is the logarithm of the ratio of book value of equity plus deferred taxes to market capitalization, with the exception that book-to-market ratio values greater than the 0.995 fractile or less than the 0.005 fractile are set to equal the 0.995 and 0.005 fractile values, respectively. DVOL is the logarithm of the dollar trading volume. PRICE is the logarithm of the reciprocal of the share price. YLD is the logarithm, of the dividend yield; RET2–3, RET4–6, RET7–12 equal the logarithms of the cumulative returns over the second through third, fourth through sixth, and seventh through 12th months prior to the current month, respectively. NYDVOL is the value of DVOL if the stock trades on NYSE/AMEX, and zero otherwise; NADVOL is the value of DVOL if the stock trades on NASdaq; and zero otherwise. The estimates in the column labeled 'Raw' are the coefficients estimated using Eqs. (8) and (9), while those in the column labeled 'Purged' are from Eqs. (8) and (10). All coefficients are multiplied by 100. t-statistics are in parentheses. The variables are measured as the deviation from the cross-sectional mean in each period. The estimates in the column labeled 'Raw' are the coefficients estimated using Eqs. (8) and (9), while those in the column labeled 'Purged' are from Eqs. (8) and (9). All coefficients are multiplied by 100. t-statistics are in parentheses.

	Excess returns	Risk-adjusted returns using the Connor–Korajczyk factors		Risk-adjusted returns using the Fama–French factors	
		Raw	Purged	Raw	Purged
Intercept	0.797	0.112	0.144	0.149	0.109
	(2.52)	(1.67)	(2.58)	(2.07)	(1.52)
NASDUM	− 0.791	− 0.842	− 0.725	− 0.797	− 0.764
	(6.69)	(6.66)	(5.90)	(6.28)	(5.84)
SIZE	0.637	0.085	0.116	0.099	0.084
	(1.08)	(1.88)	(2.58)	(2.30)	(1.95)
BM	0.235	0.189	0.181	0.120	0.122
	(4.83)	(3.91)	(3.74)	(2.71)	(2.76)
PRICE	0.195	0.165	0.151	0.108	0.012
	(1.86)	(1.96)	(1.77)	(1.17)	(0.15)
NYDVOL	− 0.118	− 0.176	− 0.185	− 0.151	− 0.162
	(2.43)	(4.67)	(5.02)	(3.89)	(4.11)
NADVOL	− 0.296	− 0.404	− 0.312	− 0.306	− 0.301
	(3.56)	(5.03)	(4.63)	(4.05)	(3.88)
YLD	0.220	1.794	2.327	0.656	0.343
	(0.13)	(1.85)	(3.39)	(0.58)	(0.30)

Table 5. Continued.

	Excess returns	Risk-adjusted returns using the Connor–Korajczyk factors		Risk-adjusted returns using the Fama–French factors	
		Raw	Purged	Raw	Purged
RET2–3	0.665	1.170	0.716	0.896	1.080
	(2.34)	(4.18)	(2.93)	(3.02)	(3.60)
RET4–6	0.790	1.067	0.852	1.005	1.203
	(3.27)	(4.00)	(3.86)	(3.80)	(4.59)
RET7–12	0.874	0.329	0.371	0.669	0.977
	(6.02)	(1.71)	(2.14)	(3.22)	(5.02)

per year, so it is possible that the NADVOL variable is playing a dual role, as a volume variable and as a dummy for NASDAQ listing. Table 5 reports the results of including a separate NASDAQ dummy.

The dummy variable is highly significant and the coefficient implies that NASDAQ stocks underperform by about 9.6% per year after adjusting for factor loadings and the non-risk firm characteristics. Moreover, with the addition of the NASDAQ dummy NADVOL becomes highly significant so that trading volume has a similar effect for Nasdaq stocks as it does for the others.

Table 6 reports the results of separate regressions for the NYSE/AMEX subsample (Panel A) and the Nasdaq subsample (Panel B).[24]

Examining the results for NYSE/AMEX subsample, we again see that the book-to-market effect is attenuated considerably (the size of the coefficient is reduced by more than 50%) and its significance is also reduced considerably, when risk-adjustment is done with the FF factors. The purged coefficient of YLD is positive and significant under the CK method of risk-adjustment. Further, the coefficient on DVOL is negative and strongly significant in all of the regressions, while the lagged return effects continue to be positive and significant. The results are in fact very similar to those in Table 5.

The results for the Nasdaq subsample are reported in Panel B. The coefficient of DVOL is again significant and negative in all the regressions. While the coefficients of the other characteristics are insignificant, they are generally of the same magnitude as found in the full sample, so that the lack of significance of those that were significant in Table 5 is likely related to the smaller sample size of Nasdaq stocks. The most striking finding is that the intercept in the

[24] The average numbers of stocks in the two subsamples are 1660 and 797, respectively.

Table 6
Fama–MacBeth regression estimates of Eq. (7) using individual security data, sample split by exchange listing (NYSE/AMEX versus Nasdaq).

Coefficient estimates are time-series averages of cross-sectional OLS regressions. The dependent variable in the first column is simply the excess return, while in the second and third columns it is the excess returns risk-adjusted using the CK factors, and in the fourth and fifth columns it is the excess returns risk-adjusted using the FF factors (Dimson beats with one lag are used in each case). The independent variables are defined as follows; SIZE represents logarithm of the market capitalization of firms in billions of dollars. BM is the logarithm of the ratio of book value of equity plus deferred taxes to market capitalization, with the exception that book-to-market ratio values greater than the 0.995 fractile or less than the 0.005 fractile are set to equal the 0.995 and 0.005 fractile values, respectively. DVOL is the logarithm of the dollar trading volume. PRICE is the logarithm of the reciprocal of the share price. YLD is the logarithm, of the dividend yield; RET2–3, RET4–6, RET7–12 equal the logarithms of the cumulative returns over the second through third, fourth through sixth, and seventh through 12th months prior to the current month, respectively. NYDVOL is the value of DVOL if the stock trades on NYSE/AMEX, and zero otherwise; NADVOL is the value of DVOL if the stock trades on Nasdaq; and zero otherwise and NASDUM equals one if the stock is listed on Nasdaq and zero otherwise. The estimates in the column labeled 'Raw' are the coefficients estimated using Eqs. (8) and (9), while those in the column labeled 'Purged' are from Eqs. (8) and (10). All coefficients are multiplied by 100. t-statistics are in parentheses. The variables are measured as the deviation from the cross-sectional mean in each period. The estimates in the column labeled 'Raw' are the coefficients estimated using Eqs. (8) and (9), while those in the column labeled 'Purged' are from Eqs. (8) and (10). In Panel A, the sample consists of an average of 1660 NYSE/AMEX stocks, while in Panel B, of an average of 797 Nasdaq stocks. All coefficients are multiplied by 100. t-statistics are in parentheses.

	Excess returns	Risk-adjusted returns using the Connor–Korajczyk factors		Risk-adjusted returns using the Fama–French factors	
		Raw	Purged	Raw	Purged
Panel A: NYSE/AMEX stocks only					
Intercept	0.777	0.053	0.068	0.083	0.067
	(2.51)	(1.18)	(1.51)	(1.72)	(1.39)
SIZE	0.064	0.079	0.107	0.087	0.074
	(1.04)	(1.81)	(2.49)	(2.02)	(1.70)
BM	0.196	0.133	0.117	0.088	0.081
	(3.85)	(2.79)	(2.45)	(2.10)	(1.92)
PRICE	0.131	0.074	0.079	− 0.001	− 0.087
	(1.22)	(0.87)	(0.91)	(0.007)	(0.91)
DVOL	− 0.144	− 0.207	− 0.215	− 0.178	− 0.189
	(2.86)	(5.56)	(6.02)	(4.63)	(4.84)
YLD	− 0.377	1.450	2.044	0.121	− 0.090
	(0.23)	(1.52)	(3.03)	(0.11)	(0.08)
RET2–3	0.949	1.451	1.241	1.173	1.372
	(3.03)	(4.97)	(4.55)	(3.67)	(4.24)

Table 6. Continued.

	Excess returns	Risk-adjusted returns using the Connor–Korajczyk factors		Risk-adjusted returns using the Fama–French factors	
		Raw	Purged	Raw	Purged
RET4–6	0.889	1.210	0.052	1.103	1.317
	(3.33)	(4.43)	(4.47)	(3.95)	(4.72)
RET7–12	0.972	0.408	0.503	0.733	1.068
	(6.35)	(2.10)	(2.85)	(3.42)	(5.38)
Panel B: Nasdaq stocks only					
Intercept	0.355	– 0.425	– 0.451	– 0.486	– 0.561
	(0.90)	(3.97)	(3.14)	(3.41)	(3.81)
SIZE	0.160	– 0.024	0.061	0.202	0.181
	(0.83)	(0.10)	(0.39)	(0.90)	(0.80)
BM	0.273	0.302	0.251	0.197	0.185
	(1.89)	(1.77)	(1.79)	(1.20)	(1.13)
PRICE	0.424	0.233	0.063	0.298	0.181
	(1.82)	(0.89)	(0.30)	(1.22)	(0.75)
DVOL	– 0.322	– 0.344	– 0.283	– 0.347	– 0.368
	(2.60)	(2.34)	(2.73)	(2.43)	(2.51)
YLD	4.24	6.38	5.79	3.80	2.77
	(0.95)	(1.24)	(1.64)	(0.81)	(0.58)
RET2–3	1.063	1.095	0.117	1.46	1.638
	(1.38)	(1.22)	(0.15)	(1.68)	(1.83)
RET4–6	0.966	0.575	0.637	1.38	1.585
	(1.64)	(0.81)	(1.16)	(2.12)	(2.38)
RET7–12	0.888	0.595	0.536	0.812	1.088
	(1.87)	(1.00)	(1.27)	(1.56)	(2.04)

risk-adjusted regressions is consistently negative and significant, ranging from *minus* 0.4% to *minus* 0.6% per month, depending on the regression. Since that the non-risk characteristics are scaled to have mean zero, this finding suggests that the average Nasdaq stock underformed relative to the risk model by 5–7% per year, which is consistent with Reinganum's (1990) finding.[25]

[25] This contrasts with FF (1993a), who find for a shorter sample period (1973–1991) that the difference between NYSE and Nasdaq returns for size sorted portfolios is not significant after risk adjustment by the FF factors. Loughran (1993) attributes most of Nasdaq stocks' underperformance to the underperformance of initial public offerings which are proportionately more important on Nasdaq.

5.2. Summary of regression results

Our results may be summarized as follows. First, we find as in earlier studies that excess returns are strongly related to SIZE and BM as well as lagged returns. The introduction of PRICE, DVOL, and YLD changes the sign of the coefficient of SIZE before and after risk adjustment,[26] and NYDVOL but not NADVOL is significant. However, when a Nasdaq dummy variable is included, NADVOL becomes significant and a very large negative effect is found to be associated with Nasdaq membership. The factor model that is used to risk-adjust the returns makes relatively little difference to the results; with the exception of the YLD coefficient in the CK regressions, which increases substantially. The most consistent finding is of a strong negative effect associated with trading volume, and introduction of this variable changes the sign of the SIZE coefficient.

While the results plainly reject the null hypothesis that returns are determined by either of these specifications of the APT (or by the corresponding specifications of 'multi-factor equilibrium models' such as Merton's ICAPM) care is required in interpreting the significant coefficients on the firm characteristics for the risk-adjusted returns, especially the YLD coefficient, whose magnitude increases quite dramatically after risk adjustment using the CK factors. One interpretation is that these significant coefficients are evidence that the risk model is mis-specified, and that the priced firm characteristics are proxying for loadings on omitted factors that are priced. It is noteworthy, however, that the significance of SIZE, BM, and the volume and lagged return variables is largely robust to the choice of risk-adjustment even though the two risk models are arrived at in quite different ways, the CK factors being taken as the principal components of returns, and the FF factors being arrived at because of their relation to economic fundamentals.[27] It seems unlikely that both risk models could be misspecified in the same way which would be required if they were to yield similar results for the non-risk characteristics' rewards. For the YLD variable, however, there is a major difference in the result depending on which risk model is used (see Table 5), in that the coefficient of YLD is significant only in the CK regressions. Therefore, while the significance of this variable may be

[26] We verified that it is the volume variables that cause the sign of the coefficient of SIZE to change. While in all the regressions without DVOL the coefficient of SIZE was negative and significant, when the DVOL variable alone was added to the regression its coefficient was negative and statistically significant, while the SIZE coefficient became positive and either insignificant or only marginally significant. In addition, the significance of the DVOL coefficient is not a result of the interaction between SIZE and DVOL in that DVOL is negative and significant even if SIZE is omitted from the regressions in Table 6.

[27] FF (1993b, p. 7) note that 'although size and book-to-market equity seem like ad hoc variables for explaining average stock returns, we have reason to expect that they proxy for common risk factors in returns. In FF (1992b) we document that size and book-to-market equity are related to economic fundamentals'.

assigned to the inadequacies of the CK factors (rather than to, say, tax effects of the type described by Brennan, 1970), it seems unlikely that a similar explanation can be given for the other significant coefficients.

For example, the consistently negative relation between returns and DVOL, and the attenuation and sign reversal of the SIZE coefficient when DVOL is included, are consistent with SIZE being a proxy for DVOL, and in turn, for a liquidity premium in asset prices. The magnitude of the DVOL effect may be assessed as follows. The standard deviation of DVOL in the NYSE/AMEX and Nasdaq subsamples is 0.938 and 0.971, respectively. The parameter estimates for the excess return regressions reported in Table 5, imply that a one standard deviation increase in DVOL causes a decrease in excess return of 0.11% per month and 0.29% per month in the NYSE/AMEX and Nasdaq stocks, respectively. These appear to be plausible magnitudes for a liquidity premium. Further it is possible that the positive SIZE effect that we observe when DVOL is included results from the correlation of SIZE with DVOL and a mis-specification of the relation between required returns and volume of trading. The BM and lagged return effects apparent even in the risk-adjusted returns defy such simple explanation. Barberis et al. (1998) and Daniel et al. (1998a, b) offer alternative explanations based on investor psychology.

6. Factor risk premia

We have seen that the two sets of factors yield similar estimates for the coefficients of the non-risk firm characteristics except for the YLD and lagged return variables. In this section we compare the two sets of factor portfolios in terms of the squared Sharpe ratios of the tangency portfolios formed from them, and the ability of each set to price the other set.

MacKinlay (1995) argues that risk-based explanations of asset pricing anomalies are bounded by the plausibility of the (squared) Sharpe ratio of the tangency portfolio that they imply. Table 7 reports the mean excess returns and squared Sharpe ratios on the five CK factor portfolios for each estimation subperiod.

Since we have no assurance that the factor rotations are the same for the two subperiods, they should be treated separately. However, we include a χ^2 test of the joint hypothesis that the mean excess returns of each factor are zero for both subperiods. Only for factors 1 and 2 can we reject the null. The table reports the each factor's squared Sharpe ratio (SSR) for each subperiod, as well as an aggregate SSR for the factor portfolios, which is the SSR of the tangency portfolio formed from the factor portfolios.[28] The largest SSR's are 0.0487 and

[28] Since the portfolio returns are orthogonal the SSR of the tangency portfolio is simply the sum of the individual portfolio SSR's. See MacKinlay (1995) and references therein.

Table 7
Excess returns and sharpe ratios for the CK and FF factors, Jan. 1963 through Dec. 1995

Mean monthly excess returns and sharpe ratios for the 5 CK for the two sub-period estimations, and the FF factors for the entire period. The five CK factors are calculated a in Connor and Korajczyk (1988), and the FF factors as in FF (1993). The Sharpe Ratio corresponding to a factor is the ratio of the mean excess return of a factor to its standard deviation. The aggregate squared Sharpe Ratio is the sum of the individual factor portfolio squared Sharpe Ratios, and is the estimated squared Sharpe Ratio of the tangency portfolio formed from the 5 (orthogonal) factor portfolios. The χ^2 statistics are for the null hypothesis that the mean excess returns for the portfolios in each of two subperiods are both equal to zero. The p-value for the combined periods in Panel A aggregates the individual sub-period F-statistics using χ^2 approximations to the F-distribution.

Panel A: factors

Factor	1	2	3	4	5	Aggregate
Period 1: *January* 1963 *to December* 1979						
Mean excess return (× 100)	0.784	– 0.865	0.788	0.767	0.484	
(t-ratio)	(1.56)	(1.72)	(1.57)	(1.52)	(0.96)	
Squared sharpe ratio	0.0123	0.0150	0.0124	0.0117	0.0046	0.0559
(p-value)						(0.06)
Period 2: *January* 1980 *to December* 1995						
Mean excess return (× 100)	1.243	0.811	– 0.015	– 0.507	0.154	
(t-ratio)	(3.06)	(1.79)	(0.03)	(1.43)	(0.35)	
Squared sharpe ratio (p-value)	0.0487	0.0167	0.0000	0.0106	0.0006	0.0767
						(0.02)
Combined periods						
Squared sharpe ratio						0.0663
(p-value)						(0.01)
Overall period: H_0 *Mean Excess*						
Return = 0						
χ^2 p-value	11.78	6.17	2.45	4.36	1.04	
	(0.00)	(0.05)	(0.29)	(0.11)	(0.59)	

Panel B: FF factors

Factor	Market	SMB	HML	Aggregate
Mean excess return (× 100)	0.480	0.248	0.441	
(t-ratio)	(2.09)	(1.66)	(3.28)	
Squared sharpe ratio (p-value)	0.0123	0.0076	0.0300	0.672
				(0.00)

0.0167 for factors 1 and 2 in the second subperiod. The estimated SSRs for the tangency portfolio are 0.056 and 0.077 for the two subperiods with p-values of 0.06 and 0.02, respectively.[29] Combining the two subperiods, the average SSR over this 33 year period is 0.066 with an approximate p-value of 0.01.[30]

The aggregate SSR for the FF factors[31] of 0.067 is only marginally higher than the average value for the CK factors. The main contributor to the aggregate SSR of the FF portfolios is the HML portfolio which alone has a SSR of 0.03, or almost three times that of the market portfolio. Thus the reward for risk implied by the FF factors is similar to that implied by the CK factors.

However, the similarity of the Sharpe ratios does not imply that the two sets of factors are economically equivalent. Table 8 reports the intercepts and R^2's from regressions of the CK factors on the FF factors and the FF factors on the CK factors.

The results are reported separately for the subperiods for which the CK factors were estimated. When the CK factors are regressed on the FF factors, factors 3 and 5 have significant intercepts in the first subperiod and factor 1 in the second subperiod; a Gibbons et al. (GRS) (1989) test is able to reject the null hypothesis that the intercepts are jointly zero at with a p-value of 0.00 in the first subperiod, but is unable to reject this hypothesis in the second subperiod. When a χ^2 approximation is used to combine the GRS F-statistics for the two sub-periods, the p-value for the whole period is also 0.00. The regression of the FF factors on the CK factors yields significant intercepts in both subperiods for the market and HML portfolios but not for the SMB portfolio. The hypothesis that the intercepts are jointly equal to zero is rejected for both subperiods and the overall period. Thus, the classical version of the APT, which does not account for frictions such as taxes or illiquidity, appears to be rejected by the data.

While there is evidence that the pricing of the CK factors by the FF factors is better than that of the FF factors by the CK factors, neither set of factors is sufficient to price the other. A possible reason for this is that the average characteristics (e.g., firm size, book-to-market ratio, and trading volume) of the securities underlying the different factors differ; without adjusting for the differences in average characteristics we should not expect either set of factors to price the other. Thus, at first sight, it is surprising to find that the CK factors do not

[29] Under the null hypothesis that the factor risk premia are jointly equal to *zero* $[(T-N)/N]$SSR is distributed central $F(N, T-N)$ where N is the number of portfolios and T is the number of time-series observations. See MacKinlay (1995).

[30] The p-value is estimated by combining the two periods using the χ^2 approximation to the F-statistic.

[31] This is given by $\boldsymbol{\mu}'\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}$ where $\boldsymbol{\mu}$ is the vector of mean excess returns on the three factor portfolios and $\boldsymbol{\Omega}$ is the variance-convariance matrix.

Table 8
Intercepts from the regressions of the estimated CK factors on the FF factors, and vice versa, Jan. 1963 through Dec. 1995

The market factor, MKT, is the excess return of the FF market portfolio. SMB is the difference between the returns of a small and a large firm portfolio; HML is the difference between the returns on a high book to market ratio portfolio and a low book to market ratio portfolio. $CK_k = (k = 1, \ldots, 5)$ denotes the CK factor portfolio returns. The GRS F-statistic is the Gibbons et al. (1989) statistic for testing the hypothesis that the intercepts from the regressions jointly equal zero. The overall χ^2 statistic aggregates the GRS-F-statistics using χ^2 approximations to the F-distribution. The intercepts are multiplied by 100. t-statistics are in parentheses.

Panel A: Intercepts from the regressions of the CK factors on the FF factors

	CK_1	CK_2	CK_3	CK_4	CK_5
July 1963 to December 1979	– 0.088	– 0.736	1.92	0.390	1.23
	(1.12)	(1.78)	(5.92)	(0.77)	(2.97)
R^2	0.98	0.38	0.62	0.06	0.38
GRS F-statistic = 15.22 (p-value = 0.00)					
January 1980 to December 1995	0.189	0.473	– 0.450	– 0.292	0.039
	(2.64)	(1.18)	(1.00)	1.26)	(0.08)
R^2	0.97	0.29			
GRS F-statistic = 1.67 (p-value = 0.14)					
Overall χ^2 statistic = 84.46 (p-value)					

Panel B: Intercepts from the regressions of the FF factors on the CK factors

	MKT	SMB	HML
July 1963 to December 1979	– 0.317	– 0.122	0.80
	(5.76)	(1.42)	(7.40)
R^2	0.97	0.87	0.66
GRS F-statistic = 26.23 (p-value) = 0.00)			
January 1980 to December 1995	0.141	– 0.059	0.438
	(2.52)	(0.70)	(2.84)
R^2	0.97	(0.70)	0.38
GRS F-statistic = 3.77 (p-value = 0.01)			
Overall χ^2 statistic = 90.01 (p-value = 0.00)			

price the FF market portfolio. However, as we have seen in Table 5, there is a significant negative return associated with trading volume, and the market portfolio is strongly weighted towards firms with higher than average trading volume. As we mentioned earlier, Shanken (1985, 1987) points out that the equilibrium APT requires that the 'true' market portfolio be well-diversified with respect to the factors. It is therefore interesting to note that while we cannot be sure the FF market portfolio is the true market portfolio, the CK factors explain 97% of the variation in the FF market factor.

In Section 4 we note that the Fama–MacBeth (1973) approach understates the standard errors of the coefficients because it neglects the effect of estimation errors in the factor loadings. Applying the results of Shanken (1992, Theorem 2),[32] consideration of this estimation error requires the variance of the estimate to be multiplied by one plus the squared Sharpe ratio of the tangency portfolio formed by the factors. The estimates of the SSRs provided in Table 7 indicate a standard error understatement of about 3.3%. Such a magnitude does not alter the basic conclusions described in Section 5.2.

7. Conclusion

In this paper, we have tested a risk-based asset pricing model against specific non-risk alternatives using data on individual securities. Using individual securities is important since, as Roll (1977) and Lo and MacKinlay (1990) have shown, the use of portfolios is problematic. We use two different specifications of the factor model that is used to adjust for risk: the principal components approach of Connor and Korajczyk (1988), and the characteristic-factor based approach of (1993b). Regardless of the method used to risk-adjust returns, we find a strong negative relation between average returns and trading volume, which is consistent with a liquidity premium in asset prices. In addition, the size and book-to-market ratio effects are strong in the CK method of risk-adjustment, while the FF factors attenuate both the magnitude and significance of these effects. There is strong evidence of return momentum both before and after risk-adjustment. Finally, Nasdaq stocks have much lower returns than the other stocks in the sample after adjusting for the effects of the firm characteristics and the factor loadings. The two sets of factors offer similar risk-return tradeoffs, but are not equivalent. In particular, neither set of factors appears to price the other, though there is evidence that the FF factors price the CK factors better than the CK factors price the FF factors.

[32] Shanken (1992) assumes conditional homoscedasticity of asset returns. Jagannathan and Wang (1998) derive the asymptotic distribution of the estimators in cross-sectional regressions of the type in Black et al. (1972) and Fama–MacBeth (1973).

References

Amihud, Y., Mendelson, H., 1986. Asset pricing and the bid-ask spread. Journal of Financial Economics 17, 223–249.

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

Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. Journal of Financial Economics 49, 307–343 (this issue).

Basu, S., 1977. The investment performance of common stocks in relation to their price-earnings ratios: test of the efficient markets hypothesis. Journal of Finance 32, 663–682.

Black, F., Jensen, M., Scholes, M., 1972. The capital asset pricing model: some empirical tests. In: Jensen, M. (Ed.), Studies in the Theory of Capital Markets. Praeger Publishers, New York.

Black, F., Scholes, M., 1974. The effects of dividend yield and dividend policy on common stock prices and returns. Journal of Financial Economics 1, 1–22.

Brennan, M.J., 1970. Taxes, market valuation, and corporate financial policy. National Tax Journal 4, 417–427.

Brennan, M.J., Subrahmanyam, A., 1995. Investment analysis and price formation in securities markets. Journal of Financial Economics 38, 361–382.

Brennan, M.J., Subrahmanyam, A., 1996. Market microstructure and asset pricing: on the compensation for illiquidity in stock returns. Journal of Financial Economics 41, 341–364.

Brennan, M.J., Chordia, T., Subrahmanyam, A., 1996. Cross-sectional determinants of expected returns. In: Modest, D. (Ed), On Finance: In Honor of Fischer Black. Oxford University Press, Cary, NC, forthcoming.

Brown, S., Weinstein, M., 1983. A new approach to testing asset pricing models: the bilinear paradigm. Journal of Finance 38, 711–743.

Campbell, J., 1996. Understanding risk and return. Journal of Political Economy 104, 298–345.

Connor, G., 1984. A unified beta pricing theory. Journal of Economic Theory 34, 13–31.

Connor, G., Korajczyk, R., 1988. Risk and return in an equilibrium APT: application of a new test methodology. Journal of Financial Economics 21, 255–290.

Connor, G., Korajczyk, R., 1993. A test for the number of factors in an approximate factor model. Journal of Finance 48, 1263–1291.

Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998a. Investor Psychology, and security market under- and over-reactions. Journal of Finance, forthcoming.

Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998b. Investor overconfidence, covariance risk, and predictors of security returns. Unpublished working paper, Northwestern University, Evanston, IL.

Daniel, K., Titman, S., 1997. Evidence on the characteristics of cross sectional variation in stock returns. Journal of Finance 52, 1–33.

Dimson, E., 1979. Risk measurement when shares are subject to infrequent trading. Journal of Financial Economics 7, 197–226.

Falkenstein, E.G., 1996. Preferences for stock characteristics as revealed by mutual fund holdings. Journal of Finance 51, 111–135.

Fama, E.F., French, K.R., 1992a. The cross section of expected stock returns. Journal of Finance 47, 427–466.

Fama, E.F., French, K.R., 1992b. The economic fundamentals of size and book-to-market equity. Unpublished working paper. University of Chicago.

Fama, E.F., French, K.R., 1993a. Differences in the risks and returns of NYSE and NASD stocks. Financial Analysts Journal 49, 37–41.

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

Fama, E.F., French, K.R., 1996. Multifactor explanations for asset pricing anomalies. Journal of Finance 51, 55–84.

Fama, E.F., MacBeth, J., 1973. Risk and return: some empirical tests. Journal of Political Economy 81, 607–636.

Ferson, W., Sarkissian, S., Simin, T., 1998. The alpha factor asset pricing model: a parable. Working paper. University of Washington, Seattle, WA.

Gibbons, M.R., 1982. Multivariate tests of financial models: a new approach. Journal of Financial Economics 10, 3–27.

Gibbons, M.R., Ross, S.A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121–1152.

Glosten, L.R., Harris, L., 1988. Estimating the components of the bid/ask spread. Journal of Financial Economics 21, 123–142.

Jagannathan, R., Wang, Z., 1998. An asymptotic theory for estimating beta-pricing models using cross-sectional regression. Working paper. Northwestern University, Evanston, IL.

Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. Journal of Finance 45, 881–898.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48, 65–92.

Jegadeesh, N., Titman, S., 1995. Overreaction, delayed reaction, and contrarian profits. Review of Financial Studies 8, 973–993.

Lakonishok, J., Shleifer, A., Vishny, R., 1994. Contrarian investment, extrapolation, and risk. Journal of Finance 49, 1541–1578.

Lehmann, B., 1990. Residual risk revisited. Journal of Econometrics 45, 71–97.

Lehmann, B., Modest, D.A., 1988. The empirical foundations of the arbitrage pricing theory. Journal of Financial Economics 21, 213–254.

Litzenberger, R., Ramaswamy, K., 1979. Dividends, short-selling restrictions, tax-induced investor clienteles and market equilibrium. Journal of Financial Economics 7, 163–196.

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

Loughran, T., 1993. NYSE vs NASDAQ returns. Journal of Financial Economics 33, 241–260.

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

Merton, R.C., 1973. An intertemporal capital asset pricing model. Econometrica 41, 867–887.

Miller, M.M., Scholes, M.S., 1978. Dividends and taxes. Journal of Financial Economics 6, 333–364.

Miller, M.M., Scholes, M.S., 1982. Dividends and taxes: some empirical evidence. Journal of Political Economy 90, 1118–1141.

Petersen, M., Fialkowski, D., 1994. Posted versus effective spreads: good prices or bad quotes. Journal of Financial Economics 35, 269–292.

Reinganum, M.R., 1990. Market microstructure and asset pricing. Journal of Financial Economics 28, 127–147.

Roll, R., 1977. A critique of the asset pricing theory's tests: on past and potential testability of theory. Journal of Financial Economics 4, 129–176.

Roll, R., Ross, S., 1980. An empirical investigation of the arbitrage pricing theory. Journal of Finance 35, 1073–1103.

Shanken, J., 1985. Multibeta CAPM or equilibrium-APT?: a reply. Journal of Finance 40, 1189–1196.

Shanken, J., 1987. Multivariate proxies and arbitrage pricing relations: living with the Roll critique. Journal of Financial Economics 18, 91–110.

Shanken, J., 1992. On the estimation of beta-pricing models. Review of Financial Studies 5, 1–33.

Stambaugh, R.F., 1982. On the exclusion of assets from tests of the two-parameter model: a sensitivity analysis. Journal of Financial Economics 10, 237–268.

Stoll, H., 1978. The pricing of dealer services: an empirical study of NASDAQ stocks. Journal of Finance 33, 1152–1173.