TESTING INDEX-BASED MODELS IN U.K STOCK RETURNS

J. Richard Davies, Jonathan Fletcher and Andrew Marshall

Key Words: Index-based models, cross-sectional R^2, model misspecification

JEL Classification: G11, G12

The authors are from the University of Strathclyde.

We are grateful to the use of the Matlab programs provided by Raymond Kan and Cesare Robotti to implement the tests in this paper and to Christopher Jones for the use of the Matlab program to compute the APT factors. We are also grateful for the comments of two anonymous reviewers.

Address correspondence to Professor J. Fletcher, Department of Accounting and Finance, University of Strathclyde, Curran Building, 100 Cathedral Street, Glasgow, G4 0LN, United Kingdom, phone: +44 (0) 141 548 4963, fax: +44 (0) 552 3547, email: j.fletcher@strath.ac.uk

TESTING INDEX-BASED MODELS IN

U.K STOCK RETURNS

ABSTRACT

We examine whether index-based models similar to Cremers, Petajisto and Zitzewitz(2012) are more effective in explaining cross-sectional U.K. stock returns than the more traditional Fama and French(1993) and Carhart(1997) factor models using the two-pass cross-sectional regression approach. We find that the seven-index model has the highest cross-sectional R^2 across all models. However the superior performance of the seven-index model relative to the Fama and French(1993) and Carhart(1997) models is not robust in the multiple model comparison tests of Kan, Robotti and Shanken(2012). For these models and a conditional version of the Fama and French(1993) model, we cannot reject the null hypothesis that these models perform as least as well as the other competing models. In contrast, the four-index model of Cremers et al(2012) performs poorly relative to the competing models. Our results suggest there is little benefit in using the seven-index model as an alternative to the Carhart(1997) model in practical applications that require the estimation of expected returns.

I Introduction

The factor models of Fama and French(1993) and Carhart(1997) have been used extensively in finance research. For instance, they are used to evaluate fund performance, estimate the cost of equity capital, and in the calculation of abnormal returns in event studies. A recent example of such an application is the study by Fama and French(2010) who use both models to evaluate U.S. mutual fund performance. However, studies by Chan, Dimmock and Lakonishok(2009) and Cremers, Petajisto and Zitzewitz(2012) highlight problems in using the Fama and French(1993) and Carhart(1997) models in fund performance. Cremers et al(2012) find that neither model can correctly assign zero performance to different passive indexes in U.S. stock returns.

Cremers et al(2012) suggest modifications to the formation of the factors in the Fama and French(1993) and Carhart(1997) models to mitigate the non-zero performance of the passive indexes. They recommend value weighting the SMB and HML factors, limiting the market index to only include U.S. common stocks, and including separate factors for the relative performance of the mid-cap stocks and value/growth factors for small, mid-cap, and large stocks. Cremers et al(2012) also propose the use of index-based models as an alternative to the Fama and French(1993) and Carhart(1997) models. The index-based models are constructed from benchmark indexes provided by Standard and Poor's and Frank Russell. Cremers et al(2012) propose a four-index model and a seven-index model which include a separate index for mid-cap stocks and a separate value/growth index for large, small, and mid-cap stocks. The attraction of these indexes is that they are often used by practitioners to evaluate the performance of managed funds.

Cremers et al(2012) find that the use of the modified Fama and French(1993) and Carhart(1997) models and the index-based models perform better than the traditional Fama and French(1993) and Carhart(1997) models in providing a lower tracking error volatility of

U.S. mutual funds and in assigning zero performance for size/value portfolios of mutual funds. The index-based models perform better than the modified Fama and French(1993) and Carhart(1997) models. Gregory, Tharyan and Christidis(2013) find that modified versions of the Fama and French(1993) and Carhart(1997) models provide a marginal improvement in performance in asset pricing tests in U.K. stock returns but they do not consider the use of index-based models.

In our study, we examine whether the index-based models of Cremers et al(2012) are more effective in explaining cross-sectional U.K. stock returns than the Fama and French(1993) and Carhart(1997) models. The index models can be implemented as easily as forming the Fama and French(1993) and Carhart(1997) models. We restrict the focus of our study to linear factor models and do not consider characteristics-based models, which are considered by Chan et al(2009). We use the two-pass cross-sectional regression method to evaluate the models, where the cross-sectional R^2 (Kandel and Stambaugh(1995)) is used to judge the performance of the models. We compare the performance of the models in terms of equality of R^2 using the pairwise and multiple model comparison tests of Kan, Robotti and Shanken(2012) that allows for potential model misspecification[1].

Alternative model comparison tests are available using the Hansen and Jagannathan(1997) distance measures (HJD) within the stochastic discount factor framework (e.g. Kan and Robotti(2009), Chen and Ludvigson(2009), Li, Xu and Zhang(2010), and Gospodinov, Kan and Robotti(2013)). Kan and Zhou(2004) point out that using the first HJD and R^2 to evaluate models can lead to a different ranking of models when the zero-beta rate is unrestricted. Kan and Zhou(2004) show that the first HJD focuses on how well the models explain the prices on the test assets and the R^2 focuses on how well the models explain the

---

[1] Ludvigson(2012) advocates the use of empirical methods in asset pricing that allow for potential model misspecification and provide formal model comparison tests.

expected returns on the test assets[2]. Kan et al(2012) argue that the R^2 metric is more relevant in evaluating models when considering applications of the model, which require estimates of expected returns such as cost of capital. We use the R^2 metric in our study as our main interest in conducting such tests is to consider whether the index-based models provide a more reliable model of expected returns than the Fama and French(1993) and Carhart(1997) models for use in practical applications that require estimates of expected returns.

We evaluate the performance of the models between July 1981 and December 2010. Our set of test assets includes 16 size/book-to-market (BM) portfolios and 9 industry portfolios. As well as the two index-based models of Cremers et al(2012), and the Fama and French(1993) and Carhart(1997) models, we include models based on the capital asset pricing model (CAPM), a five-factor arbitrage pricing theory (APT) using statistical factors based on Jones(2001), and five-factor intertemporal CAPM (ICAPM) similar to Petkova(2006) (see also Kan et al(2012)). We consider conditional versions of the CAPM and Fama and French(1993) models using the scaling approach of Cochrane(1996) but do not consider conditional versions of the other multifactor models due to concerns of the danger of the models performing well due to overfitting (Dittmar(2002))[3]. We use both Ordinary Least Squares (OLS) and Generalized Least Squares (GLS) estimation.

---

[2] A similar argument can be made between the use of the first or second HJD to evaluate models as it depends upon what application of the models is being considered (see Gospodinov et al(2012), Wang and Zhang(2012)).

[3] Overfitting arises whenever a model performs well simply due to a larger number of factors in the model. In this situation it can be difficult to find significant factor risk premiums or significant factor prices of covariance risk due to the increased noise from the larger number of factors.

There are four main findings in our paper. First, we find that the seven-index model of Cremers et al(2012) has the best performance among the set of competing models we consider in terms of the highest OLS and GLS R^2. Second, we find that the seven-index model does have a significant higher GLS R^2 (but not OLS R^2) than the Fama and French(1993) and Carhart(1997) models. However in the multiple model comparison tests, for the Fama and French(1993), Carhart(1997), seven-index, and conditional Fama and French(1993) models we cannot reject the null hypothesis that these models perform at least as well as the other models in terms of OLS or GLS R^2. Third, we find that the four-index model of Cremers et al(2012) performs poorly and provides a significant lower OLS and GLS R^2 than the Fama and French(1993), Carhart(1997), seven-index, and conditional Fama and French(1993) models. We also find that the APT and ICAPM models have mixed performance. Fourth, we find some evidence against all the models we consider as even the best performing models have high intercepts and very few significant factor risk premiums or prices of covariance risk. Our results suggest that although the seven-index model does a reasonable job in explaining U.K. stock returns, there is little benefit in using the model as an alternative to the Carhart(1997) model in practical applications that require the estimation of expected returns.

Our study has two main contributions. First, we complement the recent study of Cremers et al(2012) by examining the performance of the index-based models in a different market. The focus of our study differs from Cremers et al(2012) as we focus on asset pricing tests using the R^2 metric and also conduct formal model comparison tests. Second, we extend the prior literature of U.K. evidence that examines the performance of linear factor models. A partial list includes Antoniou, Garrett and Priestley(1998), Fletcher(2001), Al-Horani, Pope and Stark(2003), Fletcher and Kihanda(2005), Gao and Huang(2008), Gregory and Michou(2009), Florackis, Gregoriou, and Kostakis(2011), Kassimatis(2011), Hwang, Gao and Owen(2012), Gregory et al(2013), and Fletcher(2010, 2013) among others[4]. We extend this literature by comparing the performance of the index-based models of Cremers et al(2012) relative to the Fama and French(1993) and Carhart(1997) models and conducting formal model comparison tests using the R^2 metric.

The closest study to ours is Gregory et al(2013) and Fletcher(2013). Fletcher(2013) examines the performance of the index-based models relative to the Fama and French(1993) and Carhart(1997) models using the second HJD measure. The focus of Fletcher(2013) is on evaluating whether the models are good benchmark models in fund performance applications where the use of second HJD is more appropriate (Wang and Zhang(2012)). Fletcher(2013) finds that the index-based models do not outperform the Carhart(1997) model in that setting. We provide a different perspective on the relative performance of the models by using the R^2 metric. Our study differs from Gregory et al(2013) as we consider the use of index-based models rather than the modified versions of the Fama and French(1993) and Carhart(1997) models We also conduct formal model comparison tests.

The paper is organized as follows. Section II describes the research methods used in the study. Section III reports the data. Section IV presents the empirical results. The final section concludes.

II Research Method

Linear factor models, such as the CAPM, predict that there is an exact linear relation between expected returns of the N assets and the corresponding betas relative to the K factors. This relation is given by:

---

[4] There is also a large literature that examines the performance of consumption-based models in U.K. stock returns such as Hyde and Sherif(2005), Fletcher(2007), Gao and Huang(2008), and Jagannathan, Marakani, Takehara, and Wang(2012) among others. We do not pursue consumption-based models here since it would require us to use quarterly return data.

$$E(r_i) = \gamma_0 + \Sigma_{k=1}^{K} \beta_{ik}\gamma_k \quad \text{for } i=1,\ldots..,N \tag{1}$$

where $E(r_i)$ is the expected return on asset i, γ_0 is the zero-beta return, β_{ik} is the beta of asset i with respect to factor k (k=1,….,K), γ_k is the factor risk premium of factor k, and K is the number of factors in the model.

An alternative representation of equation (1) is to use the linear relation between the expected returns and the covariances between the asset returns and corresponding factors (Kan et al(2012)). This relation is given by:

$$E(r_i) = \lambda_0 + \Sigma_{k=1}^{K} cov_{ik}\lambda_k \quad \text{for } i=1,\ldots..,N \tag{2}$$

where λ_0 is the zero-beta return, cov_{ik} is the covariance between the returns of asset i and factor k, and λ_k is the price of covariance risk on factor k. Kan et al point out that the factor risk premiums and factor prices of covariance risk are linked by the following relation:

$$\lambda_1 = V_f^{-1}\gamma_1 \tag{3}$$

where λ_1 as a (K,1) vector of the factor prices of covariance risk, γ_1 as a (K,1) vector of factor risk premiums, and V_f is the (K,K) covariance matrix of the K factors.

The relation in equation (3) shows that if $\lambda_k = 0$ for a given factor, then that does not imply that $\gamma_k = 0$ or vice-versa, unless V_f is a diagonal matrix. Kan et al(2012) point out that when K>1, the interpretation of factor risk premiums can be complicated when the factors are correlated with one another since the betas on a factor depend upon the other factors in the model. Focusing on the factor risk premiums addresses the question as to whether or not the factor is priced but it does not necessarily tell us whether the factor is useful in explaining cross-sectional returns given the other factors in the model[5]. Kan et al(2012) note that if we want to address whether a factor helps explain the cross-sectional expected returns, given the

[5] See Jagannathan and Wang(1998), Cochrane(2005) and Kan and Robotti(2011) for more discussion on this issue. A solution to this problem is to estimate betas on each factor in separate single regressions

other factors in the model, we should test whether the factor has a significant price of covariance risk.

In this study, we will use both equations (1) and (2) and evaluate the different factor models using the two-pass cross-sectional regression approach pioneered by Black, Jensen and Scholes(1972) and Fama and MacBeth(1973)[6]. We focus our discussion on estimating equation (1) to conserve space and draw on the presentation of methods in Kan et al(2012). In the first stage, we estimate the betas of the N assets relative to the K factors from the time-series regressions of the N asset returns on a constant and the K factors using T time-series observations on the asset returns and K factors. Define X as a (N,K+1) matrix which equals $(1_N, \beta)$, where 1_N is a (N,1) vector of ones and β is a (N,K) matrix of betas with respect to the K factors.

In the second stage, we estimate the parameters in equation (1) to minimize the weighted sum of squared pricing errors given by:

$$(u_N - X\gamma)'W(u_N - X\gamma) \tag{4}$$

where u_N is a (N,1) vector of average returns on the N assets, γ is a (K+1,1) vector of γ_0 and γ_1, and W is a (N,N) weighting matrix. The $u_N - X\gamma$ vector are the N pricing errors of the assets. If the model is well specified, then the pricing errors are equal to zero. Different weighting matrixes can be used in (4) to estimate γ. OLS estimation uses $W=I_N$ where I_N is the (N,N) identity matrix. Weighted Least Squares (WLS) estimation uses $W=\Sigma^{-1}{}_d$ where $\Sigma^{-1}{}_d$ is the (N,N) matrix containing the diagonal terms of the residual covariance matrix from the time-series regression in the first stage and zeros on the off-diagonal terms. GLS estimation

[6] Kan and Robotti(2012) provide an excellent overview of the two-pass cross-sectional regression approach. See also Jagannathan, Skoulakis and Wang(2010).

uses $W=V_N^{-1}$ where V_N is the (N,N) sample covariance matrix of the N asset returns (Maximum Likelihood (ML) estimate)[7]. The γ vector is estimated by:

$$\gamma = (X'WX)^{-1}X'Wu_N \qquad (5)$$

This estimate is identical to Fama and MacBeth(1973), when betas are fixed in the sample period, who run cross-sectional regressions for each t (t=1,..,T) of the N asset returns on X to get γ_t and estimate γ by the time-series average of γ_t. We use both OLS and GLS estimation in our study.

A useful diagnostic test of the model is the cross-sectional R^2 (Kandel and Stambaugh(1995), Kan et al(2012), Lewellen et al(2010)), The R^2 is calculated as:

$$R^2 = 1 – (Q/Q_0) \qquad (6)$$

where $Q = e'We$, e is a (N,1) vector of N pricing errors, and $Q_0 = e_0'We_0$ where $e_0 = [I_N – 1_N(1_N'W1_N)^{-1}1_N'W]u_N$. The e_0 vector captures the deviations of the average returns of the N assets from their cross-sectional average. The R^2 lies between 0 and 1 and is linked to the weighted sum of squared pricing errors (Q), where as Q increases, the R^2 falls (Kan et al). If the model is well specified, the $R^2=1$[8]. The pricing errors and R^2 are the same whether equations (1) or (2) are used. The OLS R^2 evaluates the linear factor models on how well they explain the cross-sectional average returns of the N assets. Kandel and Stambaugh(1995) and Lewellen et al(2010) show that the GLS R^2 of a linear factor model is linked to how close the factor portfolios (mimicking portfolios) are to the mean-variance

---

[7] Most studies use Σ^{-1} as the GLS weighting matrix. Kan and Zhou(2004) show that the factor risk premiums and pricing errors are the same whether using Σ^{-1} or V_N^{-1} as the weighting matrix (see also Lewellen, Nagel and Shanken(2010)).

[8] Kan et al(2012) point out that the GLS R^2 is not the same whether we use $W = V_N^{-1}$ or $W = \Sigma^{-1}$. The use of V_N^{-1} facilitates model comparison as the W remains fixed across models.

frontier. The GLS R^2 addresses how well the model captures the available risk and return opportunities in the market (Lewellen et al(2010)).

We evaluate and compare the performance of the linear factor models using the testing framework developed in Kan et al(2012)[9]. Kan et al(2012) derive the asymptotic distribution of γ and λ, under general distributional assumptions, which allows for the models to be potentially misspecified and controls for the estimation error of beta (Shanken(1992)). We use the distribution theory to examine whether there are significant factor risk premiums ($\gamma_k \neq 0$) and significant factor prices of covariance risk ($\lambda_k \neq 0$). Kan et al(2012) also derive the asymptotic distribution of the OLS and GLS R^2. The distribution theory gives a test of whether the model is correctly specified or not ($R^2=1$), a test of whether the model has no explanatory power in cross-sectional expected returns ($R^2=0$), and also the standard error of the R^2 when the model is misspecified but has some explanatory power in cross-sectional expected returns ($0<R^2<1$). We use the distribution theory of the OLS and GLS R^2 as a specification test of each factor model.

We also use the Q_c test of Kan et al(2012) as an additional model specification test. The Q_c test of Kan et al(2012) is based on the model pricing errors and is given by $e'V(e)^+e$ where V(e) is the (N,N) covariance matrix of the model pricing errors. The $^+$ term denotes the pseudo-inverse. If the model is well specified, then $Q_c=0$. Kan et al(2012) point out that the null hypothesis of $Q_c=0$ can be tested either using an asymptotic χ^2 test or an approximate *F* test.

Kan et al(2012) develop pairwise model comparison tests using the OLS and GLS R^2. The null hypothesis is that the two models have an equal OLS or GLS R^2. The test statistic is given by:

---

[9] A fuller discussion of the results in Kan et al(2012) and the model comparison tests used in our study is available on request.

$$\text{Diff} = R^2_1 - R^2_2 \quad (7)$$

where R^2_1 and R^2_2 are the OLS or GLS R^2 for models 1 and 2. The pairwise model comparison tests are complicated as the relevant test depends upon whether the models are nested to one another or not and whether the models are well specified or not. We use the model comparison tests of Kan et a(2012) to examine whether there are significant differences in the OLS or GLS R^2 for every pair of factor models. For the nested models case, we use the weighted χ^2 test in Proposition A.5 of Kan et al(2012) and for the non-nested models case, we use the normal test on Proposition A.9 of Kan et al(2012)[10].

We also use the multiple model comparison test developed by Kan et al to examine whether a benchmark model has the highest R^2 among all models. For non-nested models, this test is based on the multivariate inequality test of Wolak(1987,1989) (Likelihood Ratio(LR) test). For nested models, Kan et al(2012) show that multiple nested model comparison tests can be adapted from the pairwise nested model comparison tests. We use the model comparison tests[11] to examine if the index-based models of Cremers et al(2012) outperform the Fama and French(1993) and Carhart(1997) models. If the index-based models outperform the Fama and French(1993) and Carhart(1997) models, we expect to find significant higher OLS and GLS R^2 for the index-based models. All of the test statistics in this study are corrected for heteroskedasticity using the method of White(1980) as in Kan et al(2012).

III Data

[10] Kan et al(2012) also propose a sequential approach to test for the equality of R^2 between two non-nested factor models, which we also consider in our study.

[11] See the related model comparison tests in Kan and Robotti(2009), Li et al(2010), and Gospodinov et al(2013) within the Hansen and Jagannathan(1997) distance measure framework.

All of the data for this study is collected from the London Share Price Database (LSPD) unless otherwise specified.

A) Test Portfolios

We evaluate the performance of the linear factor models in U.K. stock returns using the monthly returns of 16 portfolios of U.K. stocks sorted by size and BM ratio and 9 industry portfolio returns between July 1981 and December 2010. We include the 9 industry portfolios in the set of test assets following the recommendation of Lewellen et al(2010). Lewellen et al(2010) suggest expanding the set of test assets to break the tight covariance structure in the use of 25 size/BM portfolios in asset pricing tests in U.S. stock returns to improve the power of the tests to discriminate between alternative models. The size/BM and industry portfolio returns are value weighted monthly buy and hold returns. Fuller details of the construction of the size/BM portfolios and industry portfolios is included in the Appendix.

Table 1 reports summary statistics of the monthly returns (%) of the size and BM portfolios (Panels A and B) and industry portfolios (Panel C). The summary statistics include the mean and standard deviation of the monthly returns. The size/BM portfolios are sorted by size in the rows (Small to Big) and BM in the columns (Low to High).

Table 1 here

Panels A and B of Table 1 show that there is a strong value effect in the average returns of the size/BM portfolios. The High portfolio has a higher average return than the Low portfolio across the four size categories. The value effect is strongest in the smallest firms, which is similar to the pattern in Fama and French(2012). The High portfolio for the two smallest categories has a smaller volatility than the Low portfolio. There is less of a size effect in the average returns of the size/BM portfolios. The Small portfolio has a higher average return and lower volatility for the High category but the opposite is true for the other 3 BM categories. The average returns in the size/BM portfolios range between 0.758% (2/2) and 1.526% (Big/High).

The summary statistics in panel C of Table 1 show that the range in average returns between the industry portfolios is less than the size/BM portfolios. The average returns range between 0.900% (Services) and 1.379% (Resources). The cyclical consumer goods industry has the highest volatility across the industry portfolios of 6.875%.

B) Linear Factor Models

In this study, we focus on domestic linear factor models. We provide full details of the construction of the factors in the Appendix. We do not consider the important issue of whether global versions of the models do a better job in explaining cross-sectional U.K. stock returns compared to domestic factor models[12]. We also do not consider a number of recent alternative factor models that have been used in U.K. stock returns. Florackis et al(2011) develop a price-impact factor to take account of liquidity and augments this new factor to the CAPM, Fama and French(1993) and Carhart(1997) models. They find that the use of the augmented factor models can help explain the momentum effect but not the size effect in U.K. stock returns. We do not consider this model since the trading volume data is not available at the start of our sample period and only becomes available for a large number of securities in 1991. We also do not consider the two-factor model of Kassimatis(2011). Kassimatis(2011) develops a third-degree stochastic dominance (TSD) factor which captures

[12] See Lewis(2011) for a recent review of global asset pricing. Studies by Griffin(2002), Fama and French(2012), and Hou, Karolyi and Kho(2011) provide evidence of the performance of domestic or regional factor models compared to global factor models in international stock returns.

the investor preferences of risk aversion for losses and risk seeking for gains and augments the TSD factor to the CAPM. Kassimatis(2011) finds that the two factor model helps explain the momentum effect in U.K. stock returns[13]. We use the following factor models in our empirical analysis:

1. CAPM

This model is a single-factor model that uses the excess returns of the U.K. stock market index (Market) as the proxy for aggregate wealth.

2. Fama and French(1993) (FF)

The FF model is a three-factor model. The factors are the excess return on the market index and two zero-cost portfolios that capture the size (SMB) and value/growth (HML) effects in stock returns.

3. Carhart(1997)

The Carhart model is a four-factor model. The factors are the three factors in the FF model and a zero-cost portfolio that captures the momentum effect (WML) in stock returns.

4. Four-index model (4-index)

This model is a four-factor model and is motivated by the four-index model in Cremers et al(2012). Cremers et al(2012) recommend the use of index-based models to capture the size and value/growth effects in stock returns. The factors include the excess returns on the largest 100 stocks (Large), the difference in returns between small stocks and large stocks (S-L), the difference in returns between high BM stocks and low BM stocks across all companies (AHML), and WML.

5. Seven-index model (7-index)

[13] An alternative model which we do not consider here is a industry based factor model along the lines of Chou, Ho and Ko(2012).

This model is a seven-factor model and is motivated by the seven-index model in Cremers et al(2012). The factors include the excess returns on the largest 100 stocks, the difference in returns between small stocks and mid-cap stocks (S-M), the difference in returns between mid-cap stocks and large stocks (M-L), the difference in returns between high BM stocks and low BM stocks across large companies (LHML), the difference in returns between high BM stocks and low BM stocks across mid-cap companies (MHML), the difference in returns between high BM stocks and low BM stocks across small companies (SHML), and WML.

6. APT

This model is a five-factor version of the APT (Ross(1976)) using statistical factors. We construct the five factors using the heteroskedastic factor analysis (HFA) of Jones(2001)[14] which allows for missing return data. Jones(2001) builds on the asymptotic principal components approach of identifying APT factors of Connor and Korajczyk(1986). Connor and Korajczyk(1986) assume that the average idiosyncratic variance is constant over time. Jones(2001) generalizes this assumption and allows for the average residual variance to change over time (residual heteroskedasticity) when extracting the APT factors. We estimate the five factors across the whole sample period.

7. ICAPM

We use a five-factor model of the ICAPM based on Campbell(1996) following a similar approach to Petkova(2006). Campbell(1996) develops a discrete-time version of Merton's (1973) ICAPM. Relevant factors include the market portfolio return and innovations in any variables that forecast future market returns. The factors include the excess market returns and innovations in the annualized dividend yield of the market index,

[14] Bekaert, Hodrick and Zhang(2009) use the HFA approach of Jones(2001) in constructing APT models in international stock returns.

one-month Treasury Bill return, term spread, and default spread. The innovations of the four state variables come from a first order Vector Autoregression (VAR) as in Petkova(2006).

8. Conditional CAPM (Cond CAPM)

The conditional version of the CAPM follows the approach of Cochrane(1996) and Lettau and Ludvigson(2001). We assume the constant and slope coefficient in the stochastic discount formulation of the CAPM are a linear function of the lag term spread as the lagged information variable. The lag term spread has been found to be an important predictor in stock returns[15]. This specification of the conditional CAPM results in three variables, which are the lag term spread, the excess market returns, and the excess market returns multiplied by the lag term spread (scaled excess market returns).

9. Conditional Fama and French(1993) (Cond FF)

The conditional version of the Fama and French(1993) model follows the same approach as above for the Cond CAPM. The variables included in the model include lag term spread, the excess factor returns, and the scaled factor excess returns (factor excess returns multiplied by the lag term spread). We demean the lag term spread when using in the Cond CAPM and Cond FF models.

Table 2 reports summary statistics of the monthly excess returns (%) of the factors in the Carhart, 4-index, and 7-index models. The table reports the mean and standard deviation of the factor excess returns. To examine the predictive ability of the lag term spread, the table also reports the slope coefficient (*t*-statistic in parentheses), and the R^2 from the predictive regression of the factor excess returns in the FF model on a constant and the lag term spread.

---

[15] See Lettau and Ludvigson(2010) for a review of the evidence of time-series predictability in stock returns.

Table 2 here

Table 2 shows that a number of factors in the factor models have significant positive average excess returns. The WML factor has the highest average excess returns across all the factors at 0.923%, which is more than two standard errors from zero, confirming the strong momentum effect in U.K. stock returns. The average excess returns on the value/growth factors are all significantly positive at the 5% level except for the LHML factor (which is significant at the 10% level), which highlights the value effect in U.K. stock returns. The value effect is stronger in smaller firms compared to larger firms as confirmed by the pattern in the average excess returns of the SHML, MHML, and LHML factors. The average excess returns of the Market and Large factors are both significantly positive at the 10% level. All of the average excess returns in the size factors are close to zero and none are more than two standard errors from zero.

Table 2 also shows that the lag term spread only has significant predictive ability for the SMB factor among the three FF factors. There is a significant positive relation between the lag term spread and the future monthly excess returns on the SMB factor. The degree of predictability in the SMB factor excess returns using the lag term spread is small as the R^2 is only 1.55%. There is an insignificant relation between the lag term spread and the future monthly excess returns of the market index and the HML factor.

IV Empirical Results

We begin our empirical analysis by estimating the cross-sectional regression between average returns and factor betas (covariances) for each model using OLS and GLS. Table 3 reports the tests of model specification using the OLS R^2 (panel A) and the GLS R^2 (panel B). The R^2 and SE(R^2) columns are the cross-sectional R^2 and the standard error of the R^2 when $0<R^2<1$. The $p(R^2=1)$ and $p(R^2=0)$ columns are the p values of the null hypothesis that the model is correctly specified (R^2=1) and that the model has no explanatory power of cross-sectional average returns (R^2=0).

Table 3 here

Table 3 shows that there is a wide spread in the OLS and GLS R^2 across the factor models. In panel A, the OLS R^2 ranges between 0.006 (CAPM) and 0.838 (7-index) and in panel B, the GLS R^2 ranges between 0.028 (CAPM) and 0.596 (7-index). The GLS R^2 is lower for most models than the corresponding OLS R^2, which is similar to Lewellen et al(2010) who argue that the use of the GLS R^2 is a more rigorous test for a factor model to perform well in. The FF and Carhart models have a higher OLS and GLS R^2 than the 4-index model but lower than the 7-index model. The Cond CAPM and Cond FF models have a marginal increase in both the OLS and GLS R^2 relative to the unconditional versions of their models. The ICAPM model has poor performance using the OLS R^2 but better performance using the GLS R^2. The four best performing models by OLS and GLS R^2 are the 7-index, Cond FF, Carhart, and FF models. The standard errors of the R^2 in Table 3 highlight the large sampling variation in the estimated R^2 using either OLS or GLS.

Table 3 shows that the null hypothesis that the R^2=1 can be rejected at the 10% significance level for the CAPM, 4-index, ICAPM, and Cond CAPM models using the OLS R^2 and for all models except the 7-index and ICAPM models using the GLS R^2. The only model for which we cannot reject the null hypothesis of the R^2=1 is for the 7-index model using either OLS or GLS. Turning to the results of the test of the null hypothesis that R^2=0, we are unable to reject the null hypothesis at the 10% significance level for the CAPM, 4-index, APT, ICAPM, Cond CAPM, and Cond FF models using the OLS R^2 and for the CAPM, 4-index, ICAPM, Cond CAPM, and Cond FF models using the GLS R^2. These results suggest that for the CAPM, 4-index, and Cond CAPM models, these models are not only misspecified using the OLS and GLS R^2 but also have no explanatory power to capture the cross-sectional spread in average returns or the risk/return opportunities of the size/BM and industry portfolios. For the FF, Carhart, and 7-index models, we can reject the null hypothesis that the R^2=0 at the 10% level using either the OLS or GLS R^2, which suggests that these models do have some explanatory power to capture cross-sectional stock returns and the risk/return opportunities in the size/BM and industry portfolios.

Table 3 shows that the 7-index model is the only model that is well specified using either the OLS or GLS R^2. However a concern for the 7-index model and the Cond FF model (which has the second highest OLS and GLS R^2) is that they have the largest number of parameters and so could perform well due to overfitting the data. We next examine whether there are significant differences in OLS and GLS R^2 between every pair of factor models using the model comparison tests of Kan et al(2012). We also use the multiple model comparison tests to examine for each model as the benchmark model to examine whether the given benchmark model has the highest R^2 across a set of models. Table 4 reports the difference in the R^2 between two models using OLS (Panel A) and GLS (Panel B). Where the difference is negative (positive), the model in the row of the table has a lower (higher) R^2 than the model in the column of the table. Panel C of the table reports the LR test and *p* value of the multiple non-nested model comparison tests. In parentheses below for the CAPM, FF, and Cond CAPM models is the *p* value for the multiple nested model comparison tests.

Table 4 here

Panels A and B of Table 4 show that there are a number of significant rejections of the null hypothesis of equal OLS and GLS R^2 between pairs of factor models. The 4-index, 7-index, and APT models provide a significant higher OLS R^2 than the CAPM model. The FF, Carhart, and 7-index models provide a significant higher GLS R^2 than the CAPM model. The conditional version of the CAPM does little to improve the performance of the CAPM as the Cond CAPM has a significant lower OLS R^2 than all the other models except the CAPM and ICAPM. Only the 7-index model has a significant higher GLS R^2 than the Cond CAPM model. The FF, Carhart, 7-index, and Cond FF models provide a significant higher OLS and GLS R^2 than the 4-index model. The 7-index model significantly outperforms the FF and Carhart models using the GLS R^2. The ICAPM model has poor performance relative to a number of models using the OLS R^2 as has a significant lower R^2 than the FF, Carhart, 7-index, APT, and Cond FF models. However there are no models with a significant higher GLS R^2 than the ICAPM[16].

Panel C of Table 4 shows that we can reject the null hypothesis that the CAPM, 4-index, and Cond CAPM models have the highest OLS or GLS R^2 across the competing models in the set of non-nested models. We are unable to reject the null hypothesis in the nested model comparison tests for the CAPM but can for Cond CAPM model using the OLS R^2. The ICAPM model is rejected as having the highest OLS R^2 among the competing models and the APT model is rejected using the GLS R^2. For the four best performing models in Table 3, 7-index, Cond FF, Carhart, and FF, we cannot reject the null hypothesis that these models have the highest OLS or GLS R^2 across the set of models.

[16] Using the sequential approach of Kan et al(2012) leads to fewer rejections of the null hypothesis of equal OLS R^2 and GLS R^2 between models.

The results of Tables 3 and 4 provide a mixed picture of the performance of the index-based models of Cremers et al(2012) relative to the FF and Carhart models[17]. The 4-index model performs poorly relative to the FF and Carhart models but the 7-index model performs better. Although the 7-index model does have a significant higher GLS R^2 than the FF and Carhart models, the superior performance does not hold in the multiple model comparison tests. The poorer performance of the 4-index model relative to the Carhart model likely stems from the way the value/growth factor is formed in the 4-index model. The APT and ICAPM models provide mixed performance in Tables 3 and 4, which depends upon the metric used. The APT model performs well using the OLS R^2 and the ICAPM performs better using the GLS R^2. However the APT (ICAPM) are dominated by other models using the multiple model comparison tests with the GLS (OLS) R^2.

The poor performance of the ICAPM with OLS R^2 differs from Kan et al(2012) who find that the ICAPM performs well compared to other models in U.S. stock returns. Kan et al(2012) use a different set of models than the models we consider. Our results of the relative performance of the index-based models compared to the FF and Carhart models differs from Cremers et al(2012). However the focus of our study differs from Cremers et al(2012). We focus on asset pricing tests and conduct formal model comparison tests, whereas Cremers et al(2012) focus on fund performance applications. It might well be that the index-based models perform better than the FF and Carhart models in fund performance applications using U.K. managed funds.

The results in Tables 3 and 4 suggest that the 7-index, Cond FF, Carhart, and FF models are the best performing models in explaining cross-sectional stock returns or

[17] We also examine the impact of using 16 size/momentum portfolios in place of the 16 size/BM portfolios. We find no significant differences in the OLS or GLS R^2 between the factor models due to the high sampling error in the model comparison tests.

capturing the risk/return opportunities in the size/BM and industry portfolios. We next examine the performance of these four models in more detail. Tables 5 and 6 report the performance of the models using OLS (Table 5) and GLS (Table 6) estimations. For each model, the table reports the zero-beta rate and the factor risk premiums[18] and *t*-statistics in parentheses which allow for potential model misspecification. The next rows report summary statistics on model pricing errors. The summary statistics include the Root Mean Squared Error (RMSE) across the N assets, and the minimum (Min) and maximum (Max) pricing error (e_i). The $p(Q_c=0)$ cell is the p value of the null hypothesis that $Q_c=0$ using the approximate F test of Kan et al(2012).

Table 5 here

Table 6 here

Tables 5 and 6 show that there is a large positive zero-beta rate using either OLS and GLS for each model. The magnitude of the zero-beta return is considerably higher than the mean return on the one-month Treasury Bill which equals 0.609% over the sample period. Lewellen et al(2010) argue that it is important to take account of the size of the cross-sectional coefficients when evaluating linear factor models. The annualized difference between the zero-beta rates in Tables 5 and 6 and the risk-free rate all exceed 3.98%, which is too large to be due to the differences in borrowing and lending rates. This result suggests that none of the models can explain the level of expected returns (Lewellen et al(2010)).

Tables 5 and 6 show that there is only one significant factor risk premium in each of the four models, which is the same whether we use OLS or GLS. There is a significant positive factor risk premium on the HML factor in the FF and Carhart models and a

[18] The factor prices of covariance risk are available on request.

significant positive factor risk premium on the SHML factor in the 7-index model. The HML factor has a significant premium in the Cond FF model but the interpretation of risk premiums is more complicated in conditional models (e.g. Lettau and Ludvigson(2001), Gospodinov and Robotti(2013)). These factors also have a significant positive price of covariance risk using either OLS or GLS, which suggests that they make a significant contribution to the OLS and GLS R^2 in these models.

There is no significant factor risk premium or price of covariance risk on the WML factor in the Carhart and 7-index models of Tables 5 and 6. This result might stem from the use of our set of test assets. The *t*-statistics in Tables 5 and 6 allow for potential model misspecification but it has little impact on the findings compared to using the standard *t*-statistics of Jagannathan and Wang(1998). This result stems from the use of portfolio factors in the models as Kan et al(2012) show that allowing for potential model misspecification has only a marginal impact when the factors are portfolio returns. The only model in Tables 5 and 6 where it is likely to have an impact is the Cond FF model given the use of the lag term spread. However the *t*-statistics on the lag term spread and the scaled factors are also low using *t*-statistics that assume the models are correctly specified.

There is little evidence against the factor models using the Q_c test, except for the Cond FF model. The null hypothesis of $Q_c=0$ can be rejected for the Cond FF model using both OLS and GLS. The RMSE is similar between the FF, Carhart, and Cond FF models. The 7-index model has the lowest RMSE and the narrowest range in pricing errors, which is similar to the R^2 tests. Although the 7-index model has the best pricing performance in terms of the pricing errors, the better performance appears to be marginal.

Lewellen et al(2010) argue that when evaluating linear factor models, it is important to take account of the magnitude of the risk premiums[19]. Where a factor is a portfolio return, the risk premium should equal the time-series mean of the factor excess returns. Comparing the factor risk premiums in Tables 5 and 6 to the average factor excess returns in Table 2 shows that there are large differences for some of the factors. There is a large difference between the factor risk premiums of the WML factor and the average excess returns of the WML factor. There is a negative premium on the excess market returns factor in FF, Carhart, and Cond FF models but the market index has positive average excess market returns. The negative market risk premium in Tables 5 and 6 is similar to Kan et al(2012). This result stems from the large zero-beta return and the lack of sufficient variation in the market betas across the portfolios.

We examine the impact of adding the returns of the one-month Treasury Bill and the returns of the factors and scaled factors in a given factor model to the set of test assets following the recommendation of Lewellen et al(2010) in unreported tests[20]. We only consider the use of GLS as the factor risk premiums will equal the average factor excess returns. Including the Treasury Bill return and the factor excess returns has a significant impact on the performance of the models. In contrast to Table 6, all four models now have a number of factors with significant risk premiums and prices of covariance risk even after adjusting for model misspecification. We can now reject the null hypothesis that the GLS $R^2=1$ for all four models at the 10% level. We can also reject the corresponding hypothesis that the GLS $R^2=0$ for each model. There is an increase in the average mispricing and a wider range of pricing errors among the size/BM and industry portfolios for all four factor

[19] See also Lewellen and Nagel(2006) for related concerns in tests of the conditional CAPM and Consumption CAPM.

[20] Results are available on request.

models. The RMSE of each model is 0.168 (FF), 0.169 (Carhart), 0.127 (7-index), and 0.182 (Cond FF). The poorer performance of the models when including the Treasury Bill and factors in the set of test assets is similar to Lewellen et al(2010).

The results in Tables 5 and 6 and the robustness tests of Lewellen et al(2010) suggest that for even the best performing models, there is some evidence against all of the models. The factor risk premiums are very different for some factors when we estimate the risk premiums to minimize the weighted sum of squared pricing errors as we do in the cross-sectional approach compared to imposing the constraint that the factor premiums equal their mean excess returns. The downside of imposing this constraint on the factor risk premiums is to increase the mispricing in the original set of test assets.

V Conclusion

We examine whether index-based models similar to Cremers et al(2012) are more effective in explaining cross-sectional U.K. stock returns than the Fama and French(1993) and Carhart(1997) models. There are four main findings in our study. First, we find that the 7-index model has the best performance among the set of models we consider in terms of the highest OLS and GLS R^2. The 7-index model is the only model where we are unable to reject the null hypothesis that the OLS or GLS R^2=1.

Second, we find that the 7-index model does have a significant higher GLS R^2 than the FF and Carhart models but not OLS R^2. However in the multiple model comparison tests, for the FF, Carhart, 7-index, and Cond FF models, we cannot reject the null hypothesis that these models perform as least as well as the other competing models in terms of OLS or GLS R^2. This finding suggests that there is little to be gained to using the 7-index model compared to the FF and Carhart models.

Third, we find that the 4-index model performs poorly using the OLS and GLS R^2. The 4-index model has a significant lower OLS and GLS R^2 than the FF, Carhart, 7-index, and Cond FF models. Likewise the APT and ICAPM models have mixed performance and are rejected in the multiple model comparison tests using the OLS R^2 (ICAPM) and GLS R^2 (APT).

Fourth, there is some evidence against all of the models in our study. The best performing models in terms of the highest OLS and GLS R^2 do have very high intercepts and there is only one significant factor risk premium or price of covariance risk in each model. When we add the Treasury Bill and the factors in a given factor model to the set of test assets for the best performing models, as recommended by Lewellen et al(2010), the pricing performance of these model in the original set of test assets becomes poorer.

In terms of practical applications, our results would caution against the use of the CAPM, 4-index, APT, and ICAPM models in practical applications which require the use of expected return inputs in U.K. stock returns. The FF, Carhart, and 7-index models are the best models to use. There is little to be gained in using a conditional version of the FF model as we find no significant differences in either the OLS or GLS R^2 between the Cond FF and FF models. Our results would also suggest there is little benefit in using the 7-index model as an alternative to the Carhart model in practical applications that require the estimation of expected returns. This finding is similar to Fletcher(2013) who finds that there is little to be gained in using the index-based models instead of the Carhart model when evaluating the models with the second HJD metric. Our findings differ from Cremers et al(2012), although the focus of our study is different from their study. It might well be that the index-based models would perform better in fund performance applications using U.K. managed funds.

Our study uses size/BM and industry portfolios to evaluate the performance of the models. An interesting extension to our study would be to evaluate the performance of the factor models in individual stock returns along the lines of Ang, Liu and Schwarz(2010) or Chordia, Goyal and Shanken(2012). We also do not consider the role of stock characteristics and whether benchmarks based on stock characteristics give more reliable benchmark models of expected returns than linear factor models. We have only considered the use of one lagged information variable in our study of conditional factor models. A fuller examination of conditional factor models using a wide range of alternative lagged information variables is worthy of more investigation. There are a number of alternative factor models that we have not examined such as augmenting the FF and Carhart models with a liquidity factor (e.g. Nguyen and Puri(2009), Florackis et al(2011)), the two-factor model of Kassimatis(2011), or the industry-based factor model of Chou et al(2012). It would be of interest to conduct model comparison tests with these models. It would be of interest to compare the performance of the index-based models relative to non-linear asset pricing models such as the four-moment CAPM utilized by Dittmar(2002). We leave an examination of these issues to future research.

Appendix

A) Formation of the Test Assets

1. Size/BM Portfolios

We form the 16 size/BM portfolios using the following approach similar to Fama and French(2012). At the start of July each year between 1981 and 2010, all stocks on the LSPD are ranked independently by their market value at the end of June and their BM ratio at the end of the previous calendar year. We use the inverse of the price-to-book ratio, which we collect from Datastream, as the BM ratio. We exclude companies with zero market values, zero and negative BM ratios, and financials. We form four size groups based on breakpoints of 3%, 13%, and 25% of aggregate market capitalization. We form four BM groups based on quartile breakpoints from the BM ratios of the largest 90% of stocks by market value (Big stocks). We form 16 portfolios using the intersection of companies of the 4x4 sorts. We calculate the monthly buy and hold returns during the next 12 months for each portfolio. The initial weights in each portfolio are value weighted using the market value of the security at the end of June.

We make a number of corrections and exclusions to the portfolio returns which we follow across forming all the portfolios and factors. Where a security has missing return observations during the year, we assign a zero return to the missing values as in Liu and Strong(2008). A security can have missing returns if it dies during the year or faces a temporary suspension. We correct for the delisting bias of Shumway(1997) by following the approach of Dimson, Nagel and Quigley(2003). A –100% return is assigned to the death event date on LSPD where the LSPD code indicates that the death is valueless. We exclude investment trusts[21], foreign companies and secondary shares.

2. Industry Portfolios

[21] Investment trusts are equivalent to closed-end U.S. mutual funds.

We form 9 industry portfolios using the following approach. At the start of each year between 1981 and 2010, we allocate all stocks on LSPD to one of nine industry portfolios based on their industry classification in LSPD at the end of the previous year. We use the following industry sectors: resources, basic industries, general industrials, cyclical consumer goods, noncyclical consumer goods, retailers, leisure and media, services, and financials. We do not include a utilities sector due to insufficient return data and exclude the closed-end funds (investment trusts) sector. There have been a number of changes in the industry classifications in LSPD during our sample period and we use the same nine sectors throughout. We calculate the monthly buy and hold returns during the next year for each industry portfolio. The initial weights in each portfolio are value weighted using the market value of the security at the end of previous year. We exclude companies with zero market values and no industry classification.

B) Formation of Factor Models

1) Factors in the Carhart model

We construct the market index for the CAPM, FF, and Carhart models using a similar approach to Dimson and Marsh(2001). At the start of each year between 1981 and 2010, we construct a value weighted portfolio of all stocks on LSPD by their market value at the start of the year. We calculate buy and hold monthly returns during the next year. We exclude companies with a zero market value. We use the excess returns of the market index using the return on the one-month U.K. Treasury Bill as the risk-free asset, which we collect from LSPD and Datastream.

We form the SMB and HML factors in the FF and Carhart models using a similar approach to Fama and French(2012). At the start of July each year between 1981 and 2010, we rank all stocks on LSPD separately by their market value at the end of June and by their BM ratio at the end of the previous calendar year. We next form two size groups (Small and

Big) using a breakpoint of 90% by aggregate market capitalization where the Small stocks are the companies with smallest 10% by market value and the Big stocks are the companies with the largest 90% by market value. We form three BM groups (Growth, Neutral, and Value) using break points of the 30^{th} and 70^{th} percentiles of the BM ratios of Big stocks. We then construct six portfolios of securities at the intersection of the size and BM groups (SG, SN, SV, BG, BN, BV). We calculate the monthly buy and hold return for the six portfolios during the next 12 months. The initial weights are set equal to the market value weights at the end of June. We exclude companies with a zero market value, zero or negative BM ratios, and financials. The SMB factor is the difference in the average return of the three small firm portfolios (SG, SN, SV) and the average return of the three large firm portfolios (BG, BN, BV). The HML factor is the average of HML_S and HML_B where HML_S is the difference in portfolio returns of SV and SG and HML_B is the difference in portfolio returns of BV and BG. The HML_S and HML_B zero-cost portfolios capture the value effect in Small stocks and Big stocks respectively.

We form the WML factor in the Carhart model using a similar approach to Fama and French(2012). At the start of each month between July 1981 and December 2010, we rank all stocks on LSPD separately by their market value at the end of June and by their past cumulative buy and hold return during months –12 to –2. We form two size groups as for the 6 size/BM portfolios and three momentum groups (Losers, Neutral, and Winners) using break points of the 30^{th} and 70^{th} percentiles of the past cumulative returns of Big stocks. We then construct six portfolios of securities at the intersection of the size and momentum groups (SL, SN, SW, BL, BN, BW). We calculate the value weighted monthly return for the six portfolios using the market value weights at the end of the previous month. We exclude companies with a zero market value, and companies with less than 12 past return observations. The WML factor is the average of WML_S and WML_B where WML_S is the difference in portfolio returns of SW and SL and WML_B is the difference in portfolio returns of BW and BL. The WML_S and WML_B zero-cost portfolios capture the momentum effect in Small stocks and Big stocks respectively.

There are some differences in our approach to forming the factors in the Carhart model compared to Gregory et al(2013)[22]. Gregory et al(2013) use the Financial Times All Share (FTA) index as the market index, whereas we use the value weighted index of all stocks. Gregory et al(2013) only use stocks on the Main market, whereas we also include stocks on the smaller stock markets such as the Alternative Investment Market (AIM). Gregory et al(2013) form two size groups using the median market value of the largest 350 stocks and form the three BM (momentum) groups using the 30% and 70% percentiles of the largest 350 stocks. They use the largest 350 stocks as argue that it captures the investment universe of U.K. institutional investors. We define large stocks as the largest 90% by market value which is a similar approach.

2) Factors in the 4-index and 7-index models

We form the index-based models using a similar approach to Cremers et al(2012). We form the factors from two BM index portfolios, three size index portfolios, and six size/BM index portfolios. We form the index portfolios as follows. At the start of July each year between 1981 and 2010, we rank all stocks on LSPD by their market value at the end of June. We exclude stocks with the smallest 1%[23] by market value when forming the index portfolios as the Russell indexes used by Cremers et al(2012) do not include the very smallest stocks.

[22] Gregory et al(2013) also form an alternative set of factors using only the largest 350 stocks, which we do not pursue in this study.

[23] Dimson and Marsh(2001) refer to these stocks as micro-cap stocks.

We form the two BM index portfolios across all stocks. We rank all companies by their BM ratio at the end of the previous calendar year and group the top 1/3 (by BM) into a All/High portfolio and the bottom 1/3 into a All/Low portfolio. We exclude companies with a zero or negative BM ratio, and financials. We form three size index portfolios across all stocks. The first index (Large) is the portfolio of the largest 100 stocks by market value. The second index (Mid) includes the companies which are ranked 101 to the largest 90% of companies by market value, which captures the mid-cap stocks. The third index (Small) includes smallest 9% of stocks by market value. Dimson and Marsh(2001) refer to these companies as low-cap stocks.

We form six size/BM index portfolios. For each size index, we rank all stocks in the index by their BM ratio at the end of the previous calendar year. We exclude companies with a zero or negative BM ratio, and financials. We group the top 1/3 of companies (by BM ratio) into a High portfolio and the bottom 1/3 of companies into a Low portfolio. The six size/BM portfolios are Large/High, Large/Low, Mid/High, Mid/Low, Small/High, Small/Low.

For each index portfolio, we construct a value weighted portfolio and calculate monthly buy and hold monthly returns during the next 12 months. The initial weights are set to the market value weights at the end of June. We form the 4-index and 7-index models from the index portfolios. The 4-index model includes the excess returns on the Large size index (relative to the one-month Treasury Bill return), the difference in returns between the Small and Large size index portfolios (S-L), the difference in returns between the All/High and All/Low BM index portfolios (AHML), and WML. The 7-index model includes the excess returns on the Large size index, the difference in returns between the Small and Mid size index portfolios (S-M), the difference in returns between the Mid and Large size index portfolios (M-L), the difference in returns between the Large/High and Large/Low size/BM index portfolios (LHML), the difference in returns between the Mid/High and Mid/Low size/BM index portfolios (MHML), the difference in returns between the Small/High and Small/Low size/BM index portfolios (SHML), and WML.

3. APT

We construct the APT models using the HFA of Jones(2001). Jones extends the asymptotic principal components analysis of Connor and Korajczyk (1986, 1987) to allow for residual heteroskedasticity. We use the first five statistical factors from the HFA applied to the monthly stock return data of all companies on LSPD between July 1981 and December 2010. We exclude investment trusts, foreign companies, and secondary shares. We include companies with missing return data using the approach of Connor and Korajczyk(1987).

4. ICAPM

We construct the ICAPM model using a similar approach to Petkova(2006) and Kan et al(2012). The factors include the excess market returns over the one-month Treasury Bill return and innovations in four state variables. The state variables include the annualized dividend yield of the market index, one-month Treasury Bill return, term spread, and default spread. The term spread is the difference in the annualized yields on long-term U.K. government bonds and the three-month Treasury Bill. The yield on long-term government bonds is collected from the International Financial Statistics U.K. country table from the IMF. The default spread is the difference in price returns of the U.K. corporate bond and government bond indexes. Up until the end of 2006, the FT Fixed Interest Securities and FT Government Securities indexes are used as the corporate and government bond indexes as in Mouselli, Michou and Stark(2008). For the remainder of the sample period, we use the Barclays Capital Sterling Aggregate Corporate and Government bond indexes as the corporate and government bond indexes. We collect the bond index data from Datastream.

We estimate the innovations in the four state variables using a first-order VAR including the excess market returns and the state variables as in Petkova(2006). All of the variables are demeaned in the VAR. We do not orthogonalize the innovations in the state variables as in Kan et al(2012). Our approach differs from their study in that they also include the SMB and HML factors in the VAR system.

Table 1 Summary Statistics of the Test Assets

Panel A	Low	2	3	High
Mean				
Small	0.764	0.758	1.109	1.526
2	0.796	0.892	1.075	1.471
3	0.831	1.097	1.189	1.391
Big	0.846	1.119	1.233	1.129

Panel B	Low	2	3	High
Standard Deviation				
Small	6.186	5.715	5.140	4.937
2	6.045	5.319	5.452	5.589
3	6.162	5.535	5.629	6.239
Big	4.699	5.157	4.938	5.365

Panel C	Mean	Standard Deviation
Industry		
Resources	1.379	6.142
Basic industries	1.089	5.885
General industrials	1.095	5.942
Cyclical consumer goods	0.991	6.875
Noncyclical consumer goods	1.298	4.530
Retailers	1.013	5.058
Leisure and media	1.069	5.856
Services	0.900	5.805
Financials	0.994	5.767

The table includes summary statistics of the monthly returns (%) of 16 size/BM portfolios and 9 industry portfolios between July 1981 and December 2010. Panels A and B report the mean and standard deviation of the size/BM portfolio returns. The size/BM portfolios are sorted by size in the rows (Small to Big) and BM in the column (Low to High). Panel C reports the mean and standard deviation of the 9 industry portfolio returns.

Table 2 Summary Statistics of Factor Excess Returns

	Mean	Standard Deviation	Lag Term Spread	R^2
Market	0.470	4.559	0.155 (1.04)	0.003
SMB	0.006	3.151	0.239 (2.40)[1]	0.015
HML	0.512	2.637	0.046 (0.68)	0.001
WML	0.923	3.604		
Large	0.477	4.418		
S-L	-0.050	3.334		
AHML	0.393	3.089		
S-M	-0.071	2.107		
M-L	0.020	2.702		
LHML	0.323	3.689		
MHML	0.503	4.523		
SHML	0.605	3.319		

[1] Significant at 5%

The table reports summary statistics of the excess returns of the factors in the linear factor models between July 1981 and December 2010. The summary statistics include the mean and standard deviation of excess monthly returns (%). The final two columns report the slope coefficient (*t*-statistic in parentheses) and the R^2 from the predictive regression of the factor excess returns in the Fama and French(1993) model on a constant and the lagged term spread. The *t*-statistics are corrected for the effects of heteroskedasticity using the method of White(1980). Market is the excess returns on the value weighted market index. SMB, HML, and WML are zero-cost portfolios of the size, value/growth, and momentum effects in U.K. stock returns. Large is the excess returns on a value weighted portfolio of the largest 100 companies. S-L, S-M and M-L are zero-cost portfolios of the difference in returns between small companies and large companies, small companies and mid-cap companies, and between mid-cap companies and large companies. AHML, LHML, MHML, and SHML are zero-cost portfolios of the value/growth effect in all companies, large companies, mid-cap companies, and small companies.

Table 3 Cross-Sectional R^2 and Tests of Model Misspecification

Panel A	R^2	$p(R^2=1)$	$SE(R^2)$	$p(R^2=0)$
OLS				
CAPM	0.006	0.059	0.041	0.751
FF	0.651	0.524	0.178	0.074
Carhart	0.658	0.406	0.168	0.095
4-index	0.423	0.053	0.218	0.257
7-index	0.838	0.852	0.114	0.081
APT	0.642	0.352	0.201	0.126
ICAPM	0.071	0.025	0.225	0.988
Cond CAPM	0.028	0.012	0.083	0.944
Cond FF	0.667	0.280	0.171	0.145
Panel B	R^2	$p(R^2=1)$	$SE(R^2)$	$p(R^2=0)$
GLS				
CAPM	0.028	0.003	0.054	0.292
FF	0.294	0.078	0.138	0.004
Carhart	0.301	0.062	0.139	0.021
4-index	0.101	0.003	0.092	0.403
7-index	0.596	0.461	0.156	0.002
APT	0.256	0.034	0.148	0.094
ICAPM	0.223	0.196	0.196	0.548
Cond CAPM	0.125	0.015	0.121	0.326
Cond FF	0.354	0.075	0.159	0.196

The table reports the cross-sectional R^2 using OLS (panel A) and GLS (panel B) regressions for different factor models between July 1981 and December 2010. The set of test assets are the returns of 16 size/BM and 9 industry portfolios. The $p(R^2=1)$ column is the p value of the null hypothesis that the OLS R^2 or GLS $R^2=1$ and is a test of model specification. The $p(R^2=0)$ is the p value of the null hypothesis that the OLS R^2 or GLS $R^2=0$. The $SE(R^2)$ column is the standard error of the estimated R^2. The tests come from the results in Kan et al(2012). The test statistics are corrected for the effects of heteroskedasticity using the method of White(1980).

Table 4 Model Comparison Tests

Panel A **OLS**	FF	Carhart	4-index	7-index	APT	ICAPM	Cond CAPM	Cond FF
CAPM	-0.644	-0.651	-0.417[2]	-0.832[1]	-0.635[1]	-0.064	-0.022	-0.661
FF		-0.007	0.227[1]	-0.187	0.009	0.581[1]	0.623[1]	-0.016
Carhart			0.234[2]	-0.180	0.016	0.587[1]	0.629[1]	-0.009
4-index				-0.415[1]	-0.218	0.353	0.395[2]	-0.243[1]
7-index					0.196	0.768[1]	0.810[1]	0.171
APT						0.572[1]	0.614[1]	-0.025
ICAPM							0.042	-0.596[1]
Cond CAPM								-0.639[2]

Panel B **GLS**	FF	Carhart	4-index	7-index	APT	ICAPM	Cond CAPM	Cond FF
CAPM	-0.266[1]	-0.272[1]	-0.072	-0.568[1]	-0.227	-0.194	-0.096	-0.325
FF		-0.007	0.193[1]	-0.302[1]	0.038	0.071	0.169	-0.059
Carhart			0.200[1]	-0.295[1]	0.045	0.078	0.176	-0.052
4-index				-0.495[1]	-0.154	-0.122	-0.024	-0.253[1]
7-index					0.341[1]	0.374	0.472[1]	0.243
APT						0.033	0.131	-0.098
ICAPM							0.098	-0.131
Cond CAPM								-0.229

Panel C **Multiple models**	OLS – LR	OLS – p value	GLS – LR	GLS - pvalue
CAPM	44.457	0.000 (0.524)	11.603	0.002 (0.676)
FF	1.967	0.351 (0.991)	3.936	0.104 (0.845)
Carhart	2.087	0.371	3.859	0.135
4-index	5.033	0.048	11.783	0.004
7-index	0.000	0.875	0.000	0.825
APT	1.469	0.286	6.675	0.031
ICAPM	10.238	0.002	2.483	0.125
Cond CAPM	40.380	0.000 (0.068)	5.784	0.039 (0.234)
Cond FF	1.673	0.448	1.908	0.272

[1] Significant at 5%
[2] Significant at 10%

The table reports the model comparison tests of Kan et al(2012) between different linear factor models using the 16 size/BM and 9 industry portfolio returns as the set of test assets between July 1981 and December 2010. The tests examine whether the OLS R^2 (panel A) and GLS R^2 (panel B) between two models are equal to each other. The table includes the difference in the estimated R^2 between every pair of models. Where the difference is negative (positive), the model in the row of the table has a lower (higher) R^2 than the model in the column of the table. Panel C reports the multiple non-nested model comparison tests using each of the factor models as the benchmark model. The panel reports the Likelihood Ratio (LR) test and corresponding *p* value of the null hypothesis that the benchmark model performs as well as other models in terms of the OLS R^2 and GLS R^2. In parentheses below for the CAPM, FF, and Cond CAPM models are the *p* values of the nested model comparison tests. The test statistics are corrected for the effects of heteroskedasticity using the method of White(1980).

Table 5 Performance of Best Performing Factor Models: OLS

FF	γ_0	γ_{mkt}	γ_{SMB}	γ_{HML}
Estimate	1.552	-0.484	-0.070	0.516
	(4.12)[1]	(-1.07)	(-0.39)	(2.93)[1]
	$p(Q_c=0)$	RMSE	Min e_i	Max e_i
Pricing errors	0.157	0.124	-0.247	0.257

Carhart	γ_0	γ_{mkt}	γ_{SMB}	γ_{HML}	γ_{WML}
Estimate	1.465	-0.395	-0.078	0.541	0.065
	(3.41)[1]	(-0.80)	(-0.45)	(3.18)[1]	(0.12)
	$p(Q_c=0)$	RMSE	Min e_i	Max e_i	
Pricing errors	0.140	0.123	-0.269	0.256	

7-index	γ_0	γ_{large}	γ_{S-M}	γ_{M-L}	γ_{LHML}	γ_{MHML}	γ_{SMHL}	γ_{WML}
Estimate	0.941	0.159	0.020	0.002	0.246	0.432	0.920	0.424
	(1.69)[2]	(0.26)	(0.11)	(0.01)	(1.07)	(1.21)	(3.48)[1]	(0.72)
	$p(Q_c=0)$	RMSE	Min e_i	Max e_i				
Pricing errors	0.544	0.084	-0.194	0.181				

Cond FF	γ_0	γ_{mkt}	γ_{SMB}	γ_{HML}	γ_{0z}	γ_{mktz}	γ_{SMBz}	γ_{HMLz}
Estimate	1.449	-0.376	-0.072	0.499	0.191	0.004	-0.005	-0.002
	(3.48)[1]	(-0.78)	(-0.40)	(2.76)[1]	(0.27)	(0.25)	(-0.45)	(-0.29)
	$p(Q_c=0)$	RMSE	Min e_i	Max e_i				
Pricing errors	0.077	0.122	-0.237	0.263				

[1] Significant at 5%
[2] Significant at 10%

The table reports on the OLS estimation of the four best performing linear factor models between July 1981 and December 2010. The set of test assets is 16 size/BM and 9 industry portfolio returns. For each model, the first row reports the zero-beta return and factor risk premiums (γ) and the corresponding *t*-statistics in parentheses from the regression of average returns on a constant and the factor betas. The *t*-statistics allow for potential model misspecification and are based on Kan et al(2012). The second row provides summary statistics of the model pricing errors from the cross-sectional regression. The summary statistics include the Root Mean Squared Pricing Error (RMSE), the minimum (Min e_i) pricing error, maximum (Max e_i) pricing error. The $p(Q_c=0)$ is the *p* value of the null hypothesis that $Q_c=0$ and is a test of model specification based on Kan et al. All of the test statistics are corrected for the effects of heteroskedasticity using the method of White(1980).

Table 6 Performance of Best Performing Factor Models: GLS

FF	γ_0	γ_{mkt}	γ_{SMB}	γ_{HML}				
Estimate	1.563	-0.524	0.001	0.527				
	(4.25)[1]	(-1.18)	(0.00)	(3.60)[1]				
	$p(Q_c=0)$	RMSE	Min e_i	Max e_i				
Pricing errors	0.162	0.128	-0.245	0.214				
Carhart	γ_0	γ_{mkt}	γ_{SMB}	γ_{HML}	γ_{WML}			
Estimate	1.491	-0.447	-0.001	0.537	0.096			
	(3.60)[1]	(-0.93)	(-0.00)	(3.64)[1]	(0.19)			
	$p(Q_c=0)$	RMSE	Min e_i	Max e_i				
Pricing errors	0.147	0.128	-0.273	0.228				
7-index	γ_0	γ_{large}	γ_{S-M}	γ_{M-L}	γ_{LHML}	γ_{MHML}	γ_{SMHL}	γ_{WML}
Estimate	0.958	0.137	0.039	-0.019	0.290	0.309	1.033	0.182
	(1.91)[2]	(0.24)	(0.23)	(-0.11)	(1.38)	(0.89)	(4.57)[1]	(0.33)
	$p(Q_c=0)$	RMSE	Min e_i	Max e_i				
Pricing errors	0.509	0.088	-0.213	0.196				
Cond FF	γ_0	γ_{mkt}	γ_{SMB}	γ_{HML}	γ_{0z}	γ_{mktz}	γ_{SMBz}	γ_{HMLz}
Estimate	1.645	-0.570	0.007	0.528	0.485	0.012	-0.000	-0.003
	(4.01)[1]	(-1.18)	(0.04)	(3.26)[1]	(0.71)	(0.69)	(-0.02)	(-0.41)
	$p(Q_c=0)$	RMSE	Min e_i	Max e_i				
Pricing errors	0.079	0.131	-0.217	0.214				

[1] Significant at 5%
[2] Significant at 10%

The table reports on the GLS estimation of the four best performing linear factor models between July 1981 and December 2010. The set of test assets is 16 size/BM and 9 industry portfolio returns. For each model, the first row reports the zero-beta return and factor risk premiums (γ) and the corresponding *t*-statistics in parentheses from the regression of average returns on a constant and the factor betas. The *t*-statistics allow for potential model misspecification and are based on Kan et al(2012). The second row provides summary statistics of the model pricing errors from the cross-sectional regression. The summary statistics include the Root Mean Squared Pricing Error (RMSE), the minimum (Min e_i) pricing error, maximum (Max e_i) pricing error. The $p(Q_c=0)$ is the *p* value of the null hypothesis that $Q_c=0$ and is a test of model specification based on Kan et al. All of the test statistics are corrected for the effects of heteroskedasticity using the method of White(1980).

References

Al-Horani, A, Pope, P.F. and A.W. Stark, 2003, Research and development activity and expected returns in the United Kingdom, *European Finance Review*, 7, 27-46.

Ang, A. Liu, J. and K. Schwarz, 2010, Using stocks or portfolios in tests of factor models, *Working Paper*, Columbia University.

Antoniou, A., Garrett, I. and R. Priestley, 1998, Macroeconomic variables as common pervasive risk factors and the empirical content of the arbitrage pricing theory, *Journal of Empirical Finance* 5, 221-240.

Bekaert, G., Hodrick, R.J. and X. Zhang, 2009, International stock return comovements, *Journal of* Finance, 64, 2591-2626.

Black, F., Jensen, M.C. and M. Scholes, 1972, The capital asset pricing model: Some empirical findings, In Jensen, M.C. (ed), Studies in the Theory of Capital Markets, Praeger, New York.

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

Carhart, M. M., 1997. Persistence in mutual fund performance. *Journal of Finance*, 52, 57-82.

Chan, L.K.C., Dimmock, S.G. and J. Lakonishok, 2009, Benchmarking money manager performance: Issues and evidence, *Review of Financial Studies*, 22, 4553-4599.

Chen, X. and S. Ludvigson, 2009, Land of addicts? An empirical investigation of habit-based asset pricing models, *Journal of Applied Econometrics*, 24, 1057-1093.

Chordia, T. Goyal, A. and J. Shanken, 2012, Cross-sectional asset pricing with individual stocks, *Working Paper*, Emory University.

Chou, P.H., Ho, P.H. and K.C. Ko, 2012, Do industries matter in explaining stock returns and asset-pricing anomalies?, *Journal of Banking and Finance*, 36, 355-370.

Cochrane, J.H., 1996, A cross-sectional test of an investment based asset pricing model, *Journal of Political Economy*, 104, 572-621.

Cochrane, J. H., 2005, *Asset pricing: Revised Edition* (Princeton NJ: Princeton University Press).

Connor, G. and R.A. Korajczyk, 1986, Performance measurement with the Arbitrage Pricing Theory: A new framework for analysis, *Journal of Financial Economics* 15, 373-394.

Connor, G. and R.A. Korajczyk, 1987, Estimating pervasive economic factors with missing observations, *Working Paper*, Northwestern University.

Cremers, M., Petajisto, A. and E. Zitzewitz, 2012, Should benchmark indices have alpha? Revisiting performance evaluation, *Critical Finance Review*, forthcoming.

Dimson, E. and P.R. Marsh, 2001, U.K. financial market returns 1955-2000, *Journal of Business* 74, 1-31.

Dimson, P., Nagel, S. and G. Quigley, 2003, Capturing the value premium in the U.K. 1955-2001, *Financial Analysts Journal*, 5559, 35-45.

Dittmar, R.F. 2002. Nonlinear pricing kernels, kurtosis preference and cross-section of equity returns, *Journal of Finance* 57, 369-403.

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

Fama, E.F. and K.R. French, 2010, Luck versus skill in the cross section of mutual fund returns, *Journal of Finance,* 65, 1915-1947.

Fama, E.F. and K.R. French, 2012, Size, value, and momentum in international stock returns, *Journal of Financial Economics*, 105, 457-472.

Fama, E.F., and J.D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy*, 71, 607-636.

Fletcher, J, 2001, An examination of alternative factor models in UK stock returns, *Review of Quantitative Finance and Accounting*, 16, 117-130.

Fletcher, J., 2007, Can asset pricing models price idiosyncratic risk in U.K. stock returns?, *Financial Review*, 42, 507-535.

Fletcher, J., 2010, Arbitrage and the evaluation of linear factor models in U.K. stock returns, *The Financial Review*, 45, 449-468.

Fletcher, J., 2013, Benchmark models of expected returns in U.K. portfolio performance: An empirical investigation, *International Review of Economics and Finance*, forthcoming.

Fletcher, J. and J. Kihanda, 2005, An examination of alternative CAPM Based models in U.K. stock returns, *Journal of Banking and Finance*, 29, 2995-3014.

Florackis, C., Gregorious A. and A. Kostakis, 2011, Trading frequency and asset pricing on the London stock exchange: Evidence from a new price impact ratio, *Journal of Banking and Finance*, 35, 3335-3350.

Gao, P. and K.X.D. Huang, 2008, Aggregate consumption-wealth ratio and the cross-section of stock returns: Some international evidence, *Annals of Economics and Finance*, 9, 1-37.

Gospodinov, N., Kan, R. and C. Robotti, 2013, Chi-squared tests for evaluation and comparison of asset pricing models, *Journal of Econometrics*, 173, 108-125.

Gospodinov, N., Kan, R. and C. Robotti, 2012, On the Hansen-Jagannathan distance with a no-arbitrage constraint, *Working Paper*, University of Toronto.

Gospodinov, N. and C. Robotti, 2013, Asset pricing theories, models, and tests, in Baker, H.K. and M.G. Filbeck (Eds), Portfolio Theory and Management, chapter 3, Oxford University Press.

Gregory, A. and M. Michou, 2009, Industry costs of equity: UK evidence, *Journal of Business Finance and Accounting*, 36, 679-704.

Gregory, A., Tharyan, R. and A. Christidis, 2013, Constructing and testing alternative versions of the Fama-French and Carhart models in the UK, *Journal of Business Finance and Accounting*, 40, 172-214.

Griffin, J.M., 2002, Are the Fama and French factors global or country-specific?, *Review of Financial Studies*, 15, 783-803.

Hansen, L.P. and R. Jagannathan, 1997, Assessing specification errors in stochastic discount factor models, *Journal of Finance* 52, 591-607.

Hou, K., Karolyi, G.A. and B.C. Kho, 2011, What factors drive global stock returns?, *Review of Financial Studies*, 24, 2527-2574.

Hwang, T., Gao, S. and H. Owen, 2012, A two-pass study of the CAPM: Evidence from the UK stock market, *Studies in Economics and Finance*, 29, 89-104.

Hyde, S. and M. Sherif, 2005, Consumption asset pricing models: Evidence from the U.K., *The Manchester School* 73, 343-363.

Jagannathan, R., Marakani, S., Takehara, H. and Y. Wang, 2012, Calendar cycles, infrequent decisions and the cross-section of stock returns, *Management Science*, 58, 507-522.

Jagannathan, Skoulakis, G. and Z. Wang, 2010, The analysis of the cross section of security returns, in Ait-Sahalia, Y. and L. Hansesn (Eds.), Handbook of Financial Econometrics, 2, 73-134.

Jagannathan, R. and Z. Wang, 1998, An asymptotic theory for estimating beta-pricing models using cross-sectional regression, *Journal of Finance*, 53, 1285-1309.

Jones, C.S., 2001, Extracting factors from heteroskedastic asset returns, *Journal of Financial Economics*, 62, 293-325.

Kan, R. and C. Robotti, 2009, Model comparison using the Hansen-Jagannathan distance, *Review of Financial Studies*, 22, 3449-3490.

Kan, R. and C. Robotti, 2011, On the estimation of asset pricing models using univariate betas, *Economics Letters*, 110, 117-121.

Kan, R. and C. Robotti, 2012, Evaluation of asset pricing models using two-pass cross-sectional regressions, in J.C. Duan (Eds), Handbook of Computational Finance, 223-251.

Kan, R., Robotti, C. and J. Shanken, 2012, Pricing model performance and the two-pass cross-sectional regression methodology, *Journal of Finance*, forthcoming.

Kan, R. and G. Zhou, 2004, Hansen-Jagannathan distance: Geometry and exact distribution, *Working Paper*, University of Toronto.

Kandel, S., and R.F. Stambaugh, 1995, Portfolio inefficiency and the cross-section of expected returns, *Journal of Finance*, 50, 157-184.

Kassimatis, K., 2011, Risk aversion with local risk seeking and stock returns: Evidence from the UK market, *Journal of Business Finance and Accounting*, 38, 713-739.

Lettau, M. and S. Ludvigson, 2001, Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy*, 109, 1238-1287.

Lettau, M. and S. Ludvigson, 2010, Measuring and modelling variation in the risk-return tradeoff, in Ait-Sahalia, Y. and L. Hansen (Eds.), Handbook of Financial Econometrics, 1, 617-690.

Lewellen, S. and S. Nagel, 2006, The conditional CAPM does not explain asset-pricing anomalies, *Journal of Financial Economics*, 82, 289-314.

Lewellen, J., Nagel, S. and J. Shanken, 2010, A skeptical appraisal of asset-pricing tests, *Journal of Financial Economics*, 96, 175-194.

Lewis, K.K. Global asset pricing, *Annual Review of Financial Economics*, 3, 435-466.

Li, H., Xu, Y, and X. Zhang, 2010, Evaluating asset pricing models using the second Hansen-Jagannathan distance, *Journal of Financial* Economics, 97, 279-301.

Liu, W. and N. Strong, 2008, Biases in decomposing holding period portfolio returns, *Review of Financial Studies*, 21, 2243-2274.

Ludvigson, S., 2012, Advances in consumption-based asset pricing: Empirical tests, in Constantinides, G., Harris, M. and R. Stulz (Eds), Handbook of the Economics of Finance, 2, forthcoming.

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867-888.

Michou, M., Mouselli, S. and A.W. Stark, 2008, On the information content of the Fama and French factors in the UK, *Working Paper*, Manchester Business School.

Petkova, Ralitsa, 2006, Do the Fama-French factors proxy for innovations in predictive variables, *Journal of Finance*, 61, 581-612.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341-360.

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

Shumway, T., 1997, The delisting bias in CRSP data, *Journal of Finance*, 52, 327-340.

Wang, Z. and X. Zhang, 2012, Empirical evaluation of asset pricing models: Arbitrage and pricing errors in contingent claims, *Journal of Empirical Finance*, 19, 65-78.

White, H., 1980, A heteroskedasticity consistent covariance matrix estimator and a direct test of heteroskedasticity, *Econometrica*, Vol. 48, pp. 817-838.

Wolak, F.A., 1987, An exact test for multiple inequality and equality constraints in the linear regression model, *Journal of the American Statistical Association*, 82, 782-793.

Wolak, F.A., 1989, Testing inequality constraints in linear econometric models, *Journal of Econometrics*, 31, 205-235.