

**BENCHMARK MODELS OF EXPECTED RETURNS IN U.K. PORTFOLIO
PERFORMANCE: AN EMPIRICAL INVESTIGATION**

Jonathan Fletcher

Key Words: No arbitrage, index-based models, benchmark models of expected returns

JEL Classification: G11, G12

The author is from the University of Strathclyde.

I am also grateful to C. Robotti, R. Kan, and N. Gospodinov for the use of their Matlab programs to implement some of the tests in the paper. Helpful comments received from an anonymous reviewer.

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

**BENCHMARK MODELS OF EXPECTED RETURNS IN U.K. PORTFOLIO
PERFORMANCE: AN EMPIRICAL INVESTIGATION**

ABSTRACT

I use the second Hansen and Jagannathan(1997) distance measure (HJD) to examine whether index-based models similar to Cremers, Petajisto and Zitzewitz(2012) are more reliable benchmark models of expected returns than the Fama and French(1993) and Carhart(1997) models in U.K. stock returns. I use the second HJD as it is important to take account of pricing errors over possible contingent claims when considering benchmark models that are used in fund performance applications (Wang and Zhang(2012)). I find that all of the candidate benchmark models are misspecified. I find that conditional multifactor models provide significant lower second HJD compared to the unconditional factor models. I find that there is nothing to be gained in terms of significant lower second HJD in using the index-based models compared to the conditional Carhart model. My results suggest that among the models I consider, the most reliable models are the conditional Carhart model and the conditional seven-index model of Cremers et al(2012).

I Introduction

The linear factor models of Fama and French(1993) and Carhart(1997) are used extensively in the evaluation of managed fund performance and for examining the performance of trading strategies (e.g. Alwathainani(2012)). Fama and French(2010) provide comprehensive evidence on U.S. mutual fund performance using these two models as do Cuthbertson, Nitzsche and O’Sullivan(2008) for U.K. unit trusts¹². Recent studies by Chan, Dimmock and Lakonishok(2009) and Cremers, Petajisto and Zitzewitz(2012) highlight problems in using both the Fama and French and Carhart models in fund performance. Cremers et al find that different passive indexes have significant performance relative to both models which suggests that the models are unable to correctly assign zero performance to passive trading strategies with no skill.

Cremers et al(2012) propose alternative index-based models to evaluate fund performance. The index-based models are constructed from benchmark indexes provided by Standard and Poor’s and Frank Russell and provide alternative ways to capture the size and value/growth effects in stock returns. Cremers et al find that their index-based models outperform the Fama and French(1993) and Carhart(1997) models in a number of specification tests.

I examine, using U.K. stock return data, whether index-based models similar to Cremers et al(2012) provide more reliable benchmark models of expected returns compared to the Fama and French(1993) and Carhart(1997) models. My study differs from Cremers et al in that I evaluate the models using the second Hansen and Jagannathan(1997) distance measure (HJD). Hansen and Jagannathan show that the second HJD captures the minimum

¹ U.K. unit trusts are equivalent to open-end U.S. mutual funds.

² Cuthbertson, Nitzsche and O’Sullivan(2010) provide an excellent review of open-end U.S. and U.K. fund performance evidence.

distance between a candidate stochastic discount factor³ model and the set of nonnegative stochastic discount factors that correctly price a given set of test assets⁴. The second HJD penalizes models that have pricing errors in the test assets and pricing errors in potential contingent claims, such as derivative claims in the test assets.

Wang and Zhang(2012) argue that it is important to use the second HJD to evaluate benchmark models that can be used in fund performance because the payoffs of managed funds can approximate contingent claims (see Merton(1981), Dybvig and Ross(1985), and Glosten and Jagannathan(1994) among others). The payoffs of a managed fund can approximate contingent claims either by the fund investing directly in derivatives or by engaging in dynamic trading strategies (e.g. Merton). This latter case is important as many traditional U.K. open-end and closed-end funds do not invest directly in derivatives. Benchmark models that cannot price correctly contingent claims are unreliable for evaluating fund performance (see Glosten and Jagannathan, Chen and Knez(1996), and Wang and Zhang(2012) among others for more discussion).

I consider both unconditional and conditional versions of the two index-based models similar to Cremers et al(2012) and the two empirical factor models similar to Fama and French(1993) and Carhart(1997). I also include the capital asset pricing model (CAPM) as an additional benchmark model. I estimate and evaluate the models using the second HJD between January 1959 and December 2010 using the approach developed by the recent

³ See Ferson(2003) and Cochrane(2005) for excellent reviews of the stochastic discount factor approach to asset pricing. Ferson(2012) shows how the stochastic discount factor approach can be used to unify a number of important issues in fund performance.

⁴ The set of nonnegative stochastic discount factors that correctly price the set of test assets are known as admissible stochastic discount factors. Models which do not belong to this set are known as inadmissible models.

studies of Li, Xu and Zhang(2010) and Gospodinov, Kan and Robotti(2010). I compare the performance of the models by testing the equality of the squared second HJD measures between models using the pairwise and multiple model comparison tests developed by Gospodinov et al(2010,2012a).

There are three main findings in my paper. First, I find that all of the candidate benchmark models are misspecified. None of the models are able to correctly price the N payoffs and be arbitrage free at the same time. Second, I find that the conditional multifactor models provide significant lower second HJD relative to the unconditional factor models and the conditional CAPM using the excess returns and scaled excess returns of size/dividend yield (DY) portfolios and gross Treasury Bill return as the set of payoffs. Third, I find that there is nothing to be gained in using the index-based models compared to the conditional Carhart(1997) model as there are no significant differences in the second HJD between the conditional Carhart and seven-index models. The results of the paper would suggest that among the factor models I consider, the conditional Carhart model or conditional seven-index model are the most reliable models to use in evaluating U.K. fund performance.

My study contributes to the large literature which focuses on the admissibility of benchmark models in evaluating U.S. and U.K. managed fund performance. A partial list includes Fletcher(1994), Ahn, Cao and Chretien(2009), Chan et al(2009), and Cremers et al(2012) among others. I contribute to this literature by using the second HJD to evaluate the benchmark models. Related papers by Wang and Zhang(2012), Chen and Ludvigson(2009), Li et al(2010), and Gospodinov et al(2010, 2012a) in U.S. stock returns and Fletcher(2010) in U.K. stock returns use the second HJD to evaluate different asset pricing models. My study differs from these studies, in particular Fletcher, by comparing the index-based models of Cremers et al relative to the Fama and French(1993) and Carhart(1997) models.

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

II Research Method

Ross(1978), Harrison and Kreps(1979), and Hansen and Richard(1987), among others, show that if the law of one price (LOP) holds in financial markets then there exists a stochastic discount factor m_t such that:

$$E_{t-1}(m_t x_t) = q_{t-1} \quad (1)$$

where x_t is a $(P,1)$ vector of the payoffs of P primitive assets at time t , and q_{t-1} is a $(P,1)$ vector of the costs of the P primitive assets at time $t-1$. Where financial markets satisfy the no arbitrage (NA) restriction, m_t will be positive in every state of nature (Cochrane(2005)). The stochastic discount factor will only be unique if markets are complete⁵. Equation (1) states that conditional on the information available at time $t-1$, the risk-adjusted payoffs of the primitive assets at time t has costs equal to q_{t-1} . In my study, the payoffs are the gross returns of the U.K. Treasury Bill and the excess returns of test portfolios sorted by security characteristics. In this case, the q_{t-1} vector is given by $[1;0_{P-1}]$, where 0_{P-1} is a $(P-1,1)$ vector of zeros.

Taking unconditional expectations of equation (1) results in the unconditional pricing equation:

$$E(m_t x_t) = E(q_{t-1}) \quad (2)$$

The difference between the left-hand side and right-hand side of equation (2) are the pricing errors of the primitive assets. Cochrane(1996,2005) shows that the unconditional pricing equation in (2) can be used to incorporate the impact of conditioning information without

⁵ Markets are complete when investors can buy any contingent claim (see Cochrane(2005)).

having to specify a model of conditional moments⁶. Define Z_{t-1} as a $(L+1,1)$ vector of a constant and L lagged information variables at time $t-1$. The approach of Cochrane augments the payoffs of the P assets by creating scaled payoffs where each payoff is multiplied by the L lagged information variables. The scaled payoffs of a given asset i using one of the l th lagged information variables $(x_{it}Z_{l,t-1})$ are equivalent to a dynamic trading strategy that invests $Z_{l,t-1}$ each period in asset i with a cost equal to $(q_{i,t-1}Z_{l,t-1})$. The number of payoffs (N) created by this approach equals $P*(L+1)$.

Define y_t as a candidate stochastic discount factor model. The most widely used form of stochastic discount factor models are linear factor models. In this study, I consider the models based on the CAPM, the empirical factor models of Fama and French(1993) and Carhart(1997), and the index-based models of Cremers et al(2012). The stochastic discount factor of a linear factor model can be written as:

$$y_t = \gamma_0 + \sum_{k=1}^K \gamma_{0k} f_{kt} \quad (3)$$

where f_{kt} is the value of the k th factor at time t and K is the number of factors in the models. The slope coefficients (γ_{0k}) tell us whether the factors are important in pricing the N payoffs given the other factors in the model (Cochrane(2005)). The specification in equation (3) refers to unconditional versions of the models where γ_0 and γ_{0k} are constant through time. Conditional versions of the models allow the γ_0 and γ_{0k} parameters to vary through time. The standard approach is to model γ_{0t} and γ_{0kt} as a linear function of a small number of lagged information variables in the investor's information set (Hodrick and Zhang(2001)).

One of the challenges in estimating conditional factor models is that the number of parameters increase sharply as more lagged information variables are added. The increase in

⁶ Asset pricing models can be evaluated using conditional moments in equation (1) as in Nagel and Singleton(2011).

the number of parameters raises the possibility that the model performs well due to overfitting the data (Hodrick and Zhang(2001)). To reduce the problem of overfitting, I only include one lagged information variable. I model the constant γ_{0t} and slope coefficients γ_{0kt} on the K factors as a linear function of the lagged information variable defined as z_{t-1} as $\gamma_{0t} = \gamma_0 + \gamma_1 z_{t-1}$ and $\gamma_{0kt} = \gamma_{0k} + \gamma_{1k} z_{t-1}$. The conditional version of the models can be written as:

$$y_t = \gamma_0 + \gamma_1 z_{t-1} + \sum_{k=1}^K \gamma_{0k} f_{kt} + \sum_{k=1}^K \gamma_{1k} f_{kt} z_{t-1} \quad (4)$$

Hansen and Jagannathan(1997) develop a framework to evaluate candidate stochastic discount factor models for which equation (2) might not be true. They propose two distance measures in their analysis. The first HJD measures the minimum least squares distance between y_t and the set of admissible stochastic discount factors (M) that correctly price the set of N payoffs. Hansen and Jagannathan show that the first HJD is equal to the maximum pricing error of a portfolio of N payoffs with a unit norm. Some of the admissible stochastic discount factors in set M can take on negative values and so contradict the NA restriction. Admissible stochastic discount factors in M which are negative in certain states of the world can introduce arbitrage opportunities on possible contingent claims such as derivative securities. Such stochastic discount factors are inappropriate to use in evaluating fund performance as the payoffs of managed funds can approximate the payoffs of contingent claims (e.g. Merton(1981), Glosten and Jagannathan(1994)) and so will generate pricing errors on managed funds. The payoffs of funds can approximate contingent claims either by the fund directly trading in derivatives or through dynamic trading strategies.

To take account of pricing errors over contingent claims in the evaluation of candidate stochastic discount factor models, Hansen and Jagannathan(1997) develop the second HJD. The second HJD measures the minimum least squares distance between y_t and the set of nonnegative admissible stochastic discount factors (M^+) that correctly price the N payoffs. The squared second HJD (d^{+2}) is given by:

$$d^{+2} = \min (m \in M^+) E[(y_t - m_t)^2] \quad (5)$$

The maximum pricing error interpretation of the second HJD is more complicated as the maximum pricing error over all possible contingent claims depends upon the choice of m in M^+ . Hansen and Jagannathan show that the second HJD is equal to the minmax pricing error bound over all possible contingent claims with unit norm. Wang and Zhang(2012) argue that one should use the second HJD to evaluate the suitability of benchmark models for use in fund performance. The reason is that with the second HJD, models are penalized through not only having large pricing errors over the N payoffs but also large pricing errors over contingent claims.

Hansen and Jagannathan(1997) show that the second HJD can be extended to the case where the candidate stochastic discount factor model has unknown parameters (γ). The parameters in the stochastic discount factor model can be estimated to minimize the second HJD. Hansen and Jagannathan show that we can estimate the second HJD and the parameters in the stochastic discount factor by solving the following conjugate problem:

$$d^{+2} = \min_{\gamma} \max_{\lambda} (1/T) \sum_{t=1}^T \Phi_t(\gamma, \lambda) \quad (6)$$

where $\Phi_t(\gamma, \lambda) = y_t(\gamma)^2 - m_t(\gamma, \lambda)^{+2} - 2\lambda' q_{t-1}$. The x_t vector is a $(N,1)$ vector of payoffs at time t on the primitive assets, q_{t-1} is a $(N,1)$ vector of costs on the N payoffs at time $t-1$, λ is a $(N,1)$ vector of Lagrange Multipliers, $m_t(\gamma, \lambda) = y_t(\gamma) - \lambda' x_t$, and $m_t(\gamma, \lambda)^+ = \max(0, y_t(\gamma) - \lambda' x_t)$. The Lagrange Multipliers tell us which payoffs make the largest contribution to model misspecification. If a given payoff has a zero Lagrange Multiplier then that payoff makes no contribution to model misspecification.

Hansen and Jagannathan(1997) point out that $m_t(\gamma, \lambda)^+$ is the payoff of an option that is the solution to equation (6). The $m_t(\gamma, \lambda)^+$ term is a NA admissible stochastic discount factor that belongs to set M^+ . The difference between y_t and $m_t(\gamma, \lambda)^+$ term is the minimum adjustment required to make the candidate model y_t to belong to set M^+ . The stochastic

discount factor parameters, Lagrange Multipliers, and second HJD can be solved through numerical methods. For linear factor models, Gospodinov et al(2010) provide a fast iterative analytical solution.

I estimate and compare the performance of the models using the second HJD using the results developed by the recent studies of Li et al(2010) and Gospodinov et al(2010,2012a). In the Appendix, I provide fuller details of the results of these studies and the model comparison tests but here I provide a summary of the tests I use. Li et al(2010) and Gospodinov et al(2010) derive the asymptotic distribution of the stochastic discount factor model parameters (γ) and Lagrange Multipliers (λ) under the null of a misspecified model⁷. The main challenge in deriving the asymptotic distribution is the fact that $\Phi_t(\gamma, \lambda)$ is not differentiable everywhere since the second derivative does not exist when $y_t(\gamma) - \lambda'x_t = 0$. I use the distribution theory to examine whether the factors play a significant role in pricing the N payoffs ($\gamma \neq 0$) and to examine whether the individual values of λ are equal to zero.

Hansen, Heaton and Luttmer(1995) derive the asymptotic distribution of the second HJD under the null of a misspecified model. Li et al(2010) and Gospodinov et al(2010) derive the asymptotic distribution of the second HJD under the null of a correctly specified model. I use the distribution theory of the second HJD to test whether $d^+ = 0$ as a specification test of each model. The test of a zero first HJD was derived by Jagannathan and Wang(1996). Gospodinov et al show that the same approach can be used to test for a zero second HJD. Simulation evidence in Ahn and Gadarowski(2004)) suggest that the test of Jagannathan and Wang on the first HJD tends to overreject models when N is large relative to T. As an alternative specification test, I adapt the Lagrange Multiplier (LM) test in Theorem 1 of Gospodinov et al(2012a). Gospodinov et al derive the LM test, using the first HJD

⁷ Ludvigson(2012) advocates the use of empirical methods in asset pricing that allow for potential model misspecification and facilitate model comparison tests.

framework, to examine whether $\lambda=0_N$, where 0_N is a $(N,1)$ vector of zeros. Simulation evidence in Gospodinov et al suggests that the LM test has good finite sample properties and is not subject to the overrejection problem.

Li et al(2010) and Gospodinov et al(2010) develop model comparison tests of the equality of the second HJD for two stochastic discount factor models. The model comparison tests of Li et al are refined and extended by Gospodinov et al. I test the equality of the second HJD between two models using the approach in Gospodinov et al. The test statistic is given by:

$$\text{Diff} = d^+_{F^2} - d^+_{G^2} \quad (7)$$

where $d^+_{F^2}$ and $d^+_{G^2}$ are the squared second HJD for two models F and G. I use the pairwise model comparison tests to examine whether there are significant differences in the second HJD for every pair of factor models. I also adapt and use the multiple model comparison test developed by Gospodinov et al(2012a) to examine whether a benchmark model has the lowest second HJD across the competing models. I use the model comparison tests to examine if the index-based models of Cremers et al(2012) provide a significant lower second HJD than the Fama and French(1993) and Carhart(1997) models. All of the test statistics are corrected for heteroskedasticity and serial correlation using the automatic lag selection, without prewhitening, method of Newey and West(1994).

III Data

All of the data is collected from the London Share Price Database (LSPD) provided by the London Business School unless otherwise specified. Details of the construction of the primitive assets, the factors in the linear factor models, and the lagged information variable are provided in the Appendix.

A) Primitive Assets

My set of primitive assets include the monthly gross returns of the three-month U.K. Treasury Bill, monthly excess returns of eighteen portfolios sorted by size and DY, and the scaled monthly excess returns of the size/DY portfolios between January 1959 and December 2010. I include the Treasury Bill return to tie down the expected value of the stochastic discount factor models to be just below 1 (Farnsworth, Ferson, Jackson and Todd(2002), Kan and Robotti(2008)). I do not scale the gross Treasury Bill return. I use the DY to capture the value/growth effect rather than the book-to-market (BM) ratio due to data availability. Dimson, Nagel and Quigley(2003) find that the value effect is not as strong with DY but there is a high positive correlation between zero-cost portfolios formed using the DY and BM ratios⁸.

The size/DY portfolios are formed each year and are value weighted buy and hold excess returns. All securities are grouped into four portfolios by market value in ascending order (Small to Big). Within each size portfolio, all securities are further grouped into a zero DY portfolio and four portfolios by their DY in ascending order (Low to High). For most of the years of my sample period, apart from the Small size quartile, there are an insufficient number of companies to form a zero DY portfolio for each size quartile. As a result, I group all the zero DY companies in the size quartiles 2 to Big into a single zero DY portfolio, which I refer to as the Big/Zero portfolio. Table 1 reports summary statistics of the size/DY portfolio excess returns. The table includes the mean (panel A) and standard deviation (panel B) of monthly excess returns (%).

Table 1 here

⁸ Chan, Karceski, and Lakonishok(1998) also find a DY effect in U.K. stock returns.

Table 1 shows that there is a wide spread in the average excess returns across the size/DY portfolios. The average excess returns range between 0.355% (Big/Zero) and 1.200% (Small/3). There is a clear size effect in the average excess returns across all DY classifications, where the Small portfolio has a higher average excess return than the Big portfolio. There is less of a DY effect in the mean excess returns across the size classifications. The High portfolio provides a higher mean excess return than the Low portfolio across all size classifications.

To limit the size of the N payoffs relative to the number of T observations, I only include one lagged information variable in Z_{t-1} . I create scaled excess returns on the size/DY portfolios by multiplying the portfolio excess returns by the lagged information variable. I form the scaled excess returns of the size/DY portfolios by multiplying the size/DY portfolio excess returns using the lagged term spread as the information variable. The costs of the scaled excess returns are equal to zero.

To examine the predictive ability of the lag term spread, I run predictive regressions of the size/DY portfolio excess returns on a constant and the lagged term spread in unreported tests⁹. I find that the lagged term spread has significant predictive ability of the size/DY portfolio excess returns. There is a significant positive relation between the lagged term spread and future monthly excess returns for all portfolios, except the four DY portfolios in the largest size category and the Big/Zero portfolio. A Wald test rejects the null hypothesis of no predictability across all 18 portfolios. The R^2 s are all 5.6% or under, which highlights that the predictability is small in statistical terms. The regressions provide some support to using the lagged term spread to scale the size/DY portfolio excess returns.

B) Factors

I use the following factor models in our empirical analysis:

⁹ Results are available on request.

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. I use the DY to capture the value/growth effect.

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 advocate 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 (Small-Large), the difference in returns between high DY stocks and low DY stocks across all companies (HML (All)), and WML.

5. Seven-index model (7-index)

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 (Large), the difference in returns between small stocks and mid-cap stocks (Small-Mid), the difference in returns between mid-cap stocks and large stocks (Mid-Large), the difference in returns between high DY stocks and low DY stocks across large companies (HML (Large)), the difference in returns between high DY stocks and low DY stocks across mid-cap companies

(HML (Mid)), the difference in returns between high DY stocks and low DY stocks across small companies (HML (Small)) and WML.

Table 2 reports summary statistics of the factors included in the linear factor models between January 1959 and December 2010. Table 2 shows that all of the factors have positive average excess returns except the Small-Mid factor. The WML factor has the largest mean excess return at 0.632%, which is more than two standard errors from zero. This finding confirms the strong momentum effect in U.K. stock returns. The average excess returns on the Market and Large factors are also more than two standard errors from zero. All of the size factors (SMB, Small-Large, Small-Mid, and Mid-Large) have average excess returns close to zero and none are more than two standard errors from zero. The HML and HML (All) factors have positive average excess returns and both are more than two standard errors from zero. This result confirms the value effect in U.K. stock returns. The pattern in average excess returns in the HML (Large), HML (Mid), and HML (Small) factors show that the value effect is concentrated in the largest and smallest companies. The HML (Large) and HML (Small) factors both have mean excess returns more than two standard errors from zero.

Table 2 here

In unreported tests, I examine the predictive ability of the lagged term spread for the monthly factor excess returns. I find that the lagged term spread has significant predictive ability for a number of factors. I find a significant positive relation between the lagged term spread and future monthly excess returns for the SMB, HML, Small-Large, Mid-Large, and HML (Small) factors. There is a significant negative relation between the lagged term spread and future monthly excess returns for the WML factor. All of the R^2 s are below 3.1% and so the degree of predictability is small in statistical terms.

IV Empirical Results

I begin my empirical analysis by examining whether the set M^+ is non-empty for the N payoffs using the excess returns of the 18 size/DY portfolios, scaled excess returns of the size/DY portfolios, and the gross return of the three-month U.K. Treasury Bill. Gospodinov et al(2010) show that the set M^+ can be non-empty when there are in-sample arbitrage opportunities in the set of N payoffs. Gospodinov et al show that the issue of whether in-sample arbitrage opportunities exist can be checked by running a linear programming problem with constraints that $m_t \geq 0$ for $t = 1, \dots, T$, and:

$$(1/T)\sum_{t=1}^T x_t m_t = q \quad (8)$$

where q is a $(N,1)$ vector of average costs. I run this linear program for the overall period and find that M^+ is non-empty. There exist nonnegative admissible stochastic discount factors in my sample period.

I estimate the model parameters for each linear factor model to minimize the second HJD. Table 3 reports summary statistics of the fitted stochastic discount factor values for each model and the specification test using the second HJD. The $E(y)$ and σ_y columns are the mean and standard deviation of the fitted stochastic discount factor values. The $\text{Prop}(y < 0)$ column is the proportion of the fitted stochastic discount factor values below zero (%). The d^+ column reports the second HJD and the p value of the null hypothesis of a zero second HJD is in parentheses. The $p(\lambda)$ column is the p value of the LM test of Gospodinov et al(2012a), which examines if $\lambda=0_N$. The $\text{SE}(d^+)$ column is the standard error of the second HJD under the null of a misspecified model from Hansen et al(1995). The conditional versions of the models are in bold in all the tables.

Table 3 here

Table 3 shows that all of the models are rejected as being correctly specified using the second HJD. The p values of the test for a zero second HJD are all 0 and so none of the models are able to correctly price the N payoffs and be arbitrage free at the same time. The LM test likewise rejects the null hypothesis that $\lambda=0_N$. The second HJD ranges between 0.392 (conditional 7-index) and 0.496 (unconditional CAPM). All of the conditional versions of each model have a lower second HJD than the corresponding unconditional versions of the model. The three best performing models by the magnitude of the second HJD are the conditional versions of the 7-index, Carhart, and 4-index models.

All of the factor models have a sensible value for the mean fitted stochastic discount factor values at 0.994, which is important for model comparison tests (Kan and Robotti(2008)). This result stems from including the gross Treasury Bill return in the N payoffs. The conditional versions of the models are more volatile than the unconditional versions of the models. All of the models have very few negative fitted stochastic discount factor values. The conditional Carhart model has the highest proportion of negative fitted stochastic discount factor values at 2.88%. The small proportion of negative fitted stochastic discount factor values for conditional models is a feature of using the second HJD to estimate the model parameters (see Li et al(2010)).

The results in Table 3 compare the performance of the models using the point estimates of the second HJD. I next examine whether there are significant differences in the squared second HJD between every pair of factor models using the model comparison tests. I also use the multiple model comparison tests to examine, for each model as the benchmark model, whether the given benchmark model has the lowest squared second HJD across a set of models. Panel A of Table 4 reports the difference in the squared second HJD for each pair of models. Where the difference is positive (negative), the model in the row has a higher

(lower) second HJD than the model in the column. Panel B of the table reports the LR test and p value of the multiple non-nested model comparison tests of Gospodinov et al(2012a).

Table 4 here

Panel A of Table 4 shows that there are a number of significant differences in the second HJD between models. The unconditional CAPM model is the poorest performing model and has a significant higher second HJD than every other model at the 10% level, except for the unconditional 4-index model. The null hypothesis that the unconditional CAPM model has the lowest squared second HJD among the competing models is strongly rejected in the multiple non-nested model comparison tests in panel B of Table 4 and in the nested model comparison tests. Among the unconditional multifactor models, there are no significant differences in the squared second HJD between the models. However the unconditional versions of the model perform poorly relative to the conditional multifactor models. The unconditional FF, Carhart, 4-index, and 7-index models have a significant higher second HJD than all the conditional FF, Carhart, 4-index, and 7-index models. The null hypothesis using any of the unconditional models as the benchmark model of having the lowest squared second HJD across models is rejected in the multiple non-nested and nested model comparison tests. These results provide strong support to the superior pricing performance of the conditional multifactor models relative to the unconditional models when using the second HJD to estimate and evaluate models.

Among the conditional factor models, the conditional CAPM has a significant higher second HJD than the other conditional models. The null hypothesis that the conditional CAPM has the lowest squared second HJD among the competing models is rejected in the multiple non-nested and nested model comparison tests. The conditional FF model has a

significant higher second HJD than the conditional Carhart model but among the conditional Carhart, 4-index, and 7-index models there are no significant differences in the second HJD. The null hypothesis using the conditional 4-index model as the benchmark model having the lowest squared second HJD among the models is rejected at the 10% level in the multiple non-nested model comparison tests. The conditional Carhart and 7-index models have the best performance in Table 4 when evaluating the models using the second HJD. For both models, the hypothesis using either of these models as the benchmark model that they have the lowest squared second HJD across all models cannot be rejected.

The results in Tables 3 and 4 show that conditional multifactor models outperform, in terms of lower second HJD, the unconditional factor models and the conditional CAPM. To explore the performance of some of the models in more detail, Table 5 reports the model parameters (panel A) and Lagrange Multipliers (panel B) for the two best performing models (conditional Carhart and 7-index models) and the unconditional FF model as a comparison. Panel A of the table reports the stochastic discount factor parameters (γ) and t -statistics in parentheses. The t -statistics in panel A of Table 5 are computed with misspecification-robust standard errors. The Wald test examines whether the coefficients on the lag term spread and scaled factor excess returns are jointly equal to zero. This test examines the importance of conditioning information in the conditional factor models. Panel B of the table reports the Lagrange Multipliers (λ) and t -statistics in parentheses of the five payoffs with the highest absolute t -statistics, where the t -statistics are computed under the null of a correctly specified model. The Portfolio column in panel B refers to the corresponding payoff.

Table 5 here

Panel A of Table 5 shows that there are a number of significant coefficients in the three stochastic discount factor models. In the unconditional FF model, all three factors have significant slope coefficients at the 10% level. A negative slope coefficient in the unconditional model is consistent with a positive factor risk premium. In the conditional Carhart model, there is a significant negative slope coefficient on the Market, HML, and WML factors. There is also a significant negative slope coefficient on the lag term spread and the scaled SMB factor and a significant positive slope coefficient on the scaled WML factor. The Wald test rejects the null hypothesis of the slope coefficients on the lag term spread and scaled factor excess returns being jointly equal to zero. These results suggest that conditioning information has a significant impact in improving the performance of the conditional Carhart model relative to the unconditional Carhart model.

For the conditional 7-index model, there is a significant negative slope coefficient on the HML (Large), HML (Small), and WML factors. There is a significant negative slope coefficient on the lagged term spread and a significant positive slope coefficient on the scaled WML factor at the 10% level. The Wald test rejects the null hypothesis that the slope coefficients on the lag term spread and scaled factor excess returns are jointly equal to zero. This result confirms the importance of conditioning information in improving the performance of the conditional 7-index model relative to the unconditional 7-index model.

Panel B of Table 5 shows that the gross Treasury Bill return has the largest absolute t -statistic across the three models. Gospodinov et al(2012b) show that with the choice of the cost vector in my study, $q=(1,0,0,\dots,0)'$, that the Lagrange multiplier on the Treasury Bill return is not normally distributed and so cannot be used to assess statistical significance¹⁰. This result stems from the fact that under the null of a correctly specified model, the asymptotic covariance matrix of λ is singular and some linear combinations of λ are not

¹⁰ I am grateful to the reviewer pointing out this issue.

asymptotically normally distributed. The Lagrange Multipliers of the payoffs with the largest absolute t -statistics are nearly all scaled excess returns on the size/DY portfolios. The scaled excess returns of the Big/High, Big/3, and 2/High portfolios are common across the three factor models as being among the five largest absolute t -statistics.

The results in Tables 3 to 5 show that conditional multifactor models provide superior performance relative to the unconditional factor models and conditional CAPM when using the second HJD to evaluate the models. This result is consistent with the superior performance of conditional factor models in Hodrick and Zhang(2001) who use the first HJD to evaluate models. The surprising result here is that Kan and Robotti(2009) show that much of the superior performance of the conditional factor models in terms of lower first HJD is not statistically significant when using tests that allow for potential model misspecification. I do not find this pattern when using the second HJD to compare models as the test statistics are robust to potential model misspecification.

The results also suggest that the index-based models do not significantly outperform the conditional Carhart model in terms of lower second HJD. This result differs from Cremers et al(2012). The difference could stem from the use a different metric to evaluate models and I consider conditional versions of the models in addition to the unconditional versions of the models. The two best performing models are the conditional Carhart and 7-index models. The results would suggest that these two models are the most reliable benchmark models to use in evaluating fund performance among the set of models I consider. A caveat to this interpretation is that the second HJD represents one way to evaluate the pricing errors over contingent claims¹¹. Gospodinov et al(2010) show that when markets are complete, the second HJD is a lower bound on the maximum pricing error over all contingent

¹¹ An alternative approach is the minimum discrepancy method of Almeida and Garcia(2012).

claims. As a result, they argue that one model could have a higher second HJD than another model and yet have a smaller maximum pricing error over all contingent claims. In this case, the model with the lower second HJD would not be a better model to use for pricing derivatives or alternatively a more reliable benchmark model to use in evaluating fund performance. However this argument rests on the existence of complete markets.

The analysis of the paper so far has used size/DY portfolios as the set of test assets to evaluate the different factor models. In a critique of asset pricing tests, Lewellen, Nagel and Shanken(2010) argue that the use of size/book-to-market (BM) portfolios as the test assets has limited power to discriminate between alternative models due to the tight covariance structure in the size/BM portfolios. A similar issue could arise in the use of the size/DY portfolios. Lewellen et al advocate expanding the set of test assets to break the tight covariance structure in the size/BM portfolios. I examine how sensitive my results are to using a different set of N payoffs to evaluate models. I use the excess returns of 16 size/beta portfolios¹², scaled excess returns of the size/beta portfolios, and the gross Treasury Bill return as the N payoffs. The size/beta portfolios are available between January 1961 and December 2010. Details on the construction of the size/beta portfolios are included in the Appendix.

I examine whether M^+ is non-empty in this new set of N payoffs and find that M^+ is non-empty and admissible stochastic discount factors do exist. I repeat the tests of Tables 3 and 4 using the new set of N payoffs. Table 6 reports summary statistics of the fitted stochastic discount factor values for each model and the specification tests using the second HJD and the LM test. I do not report the model comparison tests (results are available on request) but will discuss in the text.

¹² Kan, Robotti and Shanken(2012) use 25 size/beta portfolios as an alternative set of test assets when using the cross-sectional R^2 to evaluate linear factor models.

Table 6 here

Table 6 shows that the use of a different set of N payoffs has only a marginal impact on the summary statistics of fitted stochastic discount factor values and the relative performance of the models. There is an increase in the volatility of the fitted stochastic discount factor values for the unconditional Carhart, 4-index, and 7-index models and the conditional Carhart, 4-index, and 7-index models. There is only a small proportion of negative fitted stochastic discount factor values, with the conditional 7-index model having the highest proportion at 3.666%. The second HJD in Table 6 is lower than that in Table 3 for each of the models but the relative performance across the models is similar. The null hypothesis of a zero second HJD is rejected for each model as is the null hypothesis that $\lambda=0_N$, which suggests that each model is misspecified. For each factor model, the conditional version of the model provides a lower second HJD than the unconditional version of the model. The three best performing models in terms of the lowest second HJD are the conditional 7-index, Carhart, and 4-index models, which is the same as for the size/DY portfolios.

The model comparison tests using the size/beta portfolios provide a similar picture to Table 4. The main difference is that the conditional FF model does not perform as well in the size/beta portfolios. This result is due to the fact that whereas for the other conditional multifactor models, there is an increase in volatility of the fitted stochastic discount factor values using the size/beta portfolios, there is a decrease in volatility for the conditional FF model. The conditional Carhart, 4-index, and 7-index models provide a significant lower second HJD than all the unconditional models and the conditional CAPM and FF models. In the multiple non-nested and nested model comparison tests, for all the unconditional models

and the conditional CAPM and FF models the null hypothesis using these models as the benchmark that they provide the lowest squared second HJD across the models can be rejected. Among the conditional Carhart, 4-index, and 7-index models there are no significant differences in the second HJD and using the multiple model comparison tests, the hypothesis cannot be rejected for any of these models that they provide the best performance in terms of the smallest squared second HJD.

The results in Table 6 and the model comparison tests suggest that using an alternative set of N payoffs has little impact on my findings. The superior performance of the conditional multifactor models, with the exception of the conditional FF model, is robust in the different set of N payoffs and the conditional index-based models are not able to significantly outperform the conditional Carhart model. My analysis so far has used the DY when forming the value/growth factors in the Fama and French(1993), Carhart(1997), and the index-based models of Cremers et al(2012). I next examine whether my results are sensitive to the use of DY rather than the BM ratio. I collect the price-to-book ratio from Datastream and form the factor models using the BM ratio between July 1981 and December 2010. I also construct 16 size/BM portfolios and use the excess returns of the size/BM portfolios, scaled excess returns of the size/BM portfolios, and the gross Treasury Bill return as an alternative set of N payoffs. Details on the construction of the size/BM portfolios and the factor models using the BM ratio are included in the Appendix. I examine whether M^+ is non-empty using the subperiod data and find that admissible stochastic discount factors do exist.

I repeat the results of Tables 3 and 4 over the July 1981 and December 2010 subperiod. Table 7 reports summary statistics of the fitted stochastic discount factor values for each model and the specification tests using the second HJD and the LM test. I do not report the model comparison tests (results are available on request) but will discuss in the text.

Table 7 here

Table 7 shows that using the BM ratio to form the N payoffs and the value/growth factors during the July 1981 and December 2010 subperiod has some impact on the relative performance of the linear factor models. The summary statistics of the fitted stochastic discount factor values are in the main similar to Table 3 with very few negative fitted stochastic discount factor values. There is an increase in the volatility of the fitted stochastic discount factor values for the unconditional 7-index model. This increase in volatility leads to the unconditional 7-index model having the third lowest second HJD across the models. The unconditional and conditional 4-index models have poorer performance in Table 7 compared to Tables 3 and 6. The two best performing models continue to be the conditional 7-index and Carhart models with the lowest second HJD. All of the models remain misspecified as the null hypotheses of a zero second HJD and that $\lambda=0_N$ can be rejected for every model. The standard errors of the second HJD in Table 7 are higher than those in Tables 3 and 6, which reflects the use of a smaller T.

In the model comparison tests using the size/BM portfolios, there are fewer significant differences in the second HJD due to the larger sampling variation that arises due to the smaller T. The dominance of the conditional multifactor models is less apparent due to the better performance of the unconditional 7-index model and the poorer performance of the conditional 4-index model. The conditional Carhart and 7-index models continue to outperform most of the alternative models. The conditional Carhart model provides a significant lower second HJD than the unconditional CAPM, FF, Carhart, 4-index, and conditional CAPM models. The conditional 7-index model provides a significant lower second HJD than the unconditional CAPM, FF, Carhart, 4-index, and conditional CAPM, FF,

and 4-index models. In the multiple non-nested and nested model comparison tests, for the unconditional CAPM, FF, Carhart, and 4-index models, and the conditional CAPM and 4-index models the null hypothesis that using these models as the benchmark that they provide the lowest squared second HJD across models can be rejected. For the unconditional 7-index, conditional FF, Carhart, and 7-index models, I cannot reject the hypothesis that these models perform as least as well as the competing models.

The results in Table 7 and model comparison tests suggest using the BM ratio in the factor models and N payoffs has some impact on my results. However, the two best performing models continue to be the conditional Carhart and 7-index models and there is nothing to be gained in using the index-based models rather than the conditional Carhart model. The main impact is that the unconditional 7-index model performs a lot better and the unconditional and conditional 4-index models have poorer performance and the conditional multifactor models are less dominant in the subperiod.

V Conclusions

I use the second HJD to examine whether index-based models similar to Cremers et al(2012) are more reliable benchmark models of expected returns than the Fama and French(1993) and Carhart(1997) models. There are three main findings in my study. First, I find that all the benchmark models I consider are misspecified. None of the models are able to correctly price the N payoffs and be arbitrage free at the same time. None of the models are NA admissible stochastic discount factors which is important for fund performance applications (Glosten and Jagannathan(1994), Chen and Knez(1996), and Wang and Zhang(2012) among others).

Second, I find that conditional multifactor models provide significant lower second HJD than the conditional CAPM and unconditional factor models when using the excess returns and scaled excess returns of the size/DY portfolios and gross Treasury Bill return as

the set of N payoffs. This result supports the superior performance of conditional models in Hodrick and Zhang(2001) and Lettau and Ludvigson(2001) in U.S. stock returns and Fletcher and Hillier, and Fletcher(2010) in U.K. stock returns. This result is interesting in that it suggests that conditional multifactor models outperform unconditional models using the second HJD even when adjusting for potential model misspecification in the test statistics. This result differs from Kan and Robotti(2009) who find using the first HJD that there are a few significant differences between the factor models considered in Hodrick and Zhang(2001) when using test statistics that are robust to potential model misspecification.

The superior performance of some of the conditional multifactor models relative to the unconditional models is sensitive to the use of alternative payoffs and the use of the BM ratio in the July 1981 and December 2010 subperiod. The conditional FF model does not perform so well using the size/beta portfolios. The conditional 4-index model has poorer performance when using the BM ratio to form the value/growth factors and the set of payoffs in the July 1981 and December 2010 subperiod. The unconditional 7-index model performs well in the July 1981 and December 2010 subperiod relative to the other models and none of the conditional multifactor models are able to significantly outperform the unconditional 7-index model in terms of a lower second HJD.

Third, I find that the two best performing models are the conditional Carhart and 7-index models among the set of models I consider. There is no significant difference between the second HJD of the two models. This result suggests that there is nothing to be gained in using the index-based models relative to the conditional Carhart model. This result differs from Cremers et al(2012).

The results of the paper suggest that conditional versions of the Carhart or 7-index models are the most reliable benchmark models to use in evaluating fund performance among the models I consider. There is benefit in using conditional versions of the models rather

than unconditional versions of the models albeit with the caveat in Gospodinov et al(2010) that the second HJD is a lower bound on the maximum pricing error over all contingent claims in complete markets. Given that neither model is an admissible stochastic discount factor in the sample period and N payoffs used, it would be interesting to explore the use of alternative lagged information variables in the conditional models. Given the large number of lagged information variables available, one solution would be to use dynamic factor analysis as in Ludvigson and Ng(2007) to capture conditioning information as a small number of common factors from a large number of lagged information variables.

My study has used the second HJD to estimate and compare alternative models, which as Wang and Zhang(2012) argue is relevant when considering fund performance applications. For other applications, where the focus is on how well the models price the N payoffs, the use of the first HJD is more appropriate. An interesting extension to this study would be to examine how well the index-based models perform relative to the Fama and French(1993) and Carhart(1997) models using the first HJD or the cross-sectional regression R^2 in U.K. stock returns. Davis, Fletcher and Marshall(2012) provide empirical evidence on the models using the cross-sectional R^2 . I leave a fuller examination of these issues to future research.

Appendix

A) Evaluating Linear Factor Models using the Second HJD

This subsection provides a more detailed overview of the empirical methods used in the paper. I use the same notation as in Gospodinov et al(2010). I define θ as a $(KK+N,1)$ vector of model parameters (γ) and Lagrange Multipliers (λ) for a given model, where KK is the number of parameters in the stochastic discount factor model, N is the number of payoffs, and θ^* are the true values of the parameters. The M matrix is a $(KK+N, KK+N)$ matrix equal to $\lim(T \rightarrow \infty) \text{Var}[(1/\sqrt{T})\Sigma_{t=1}^T \partial\Phi_t(\theta^*)/\partial\theta]$ and H is a $(KK+N, KK+N)$ matrix given by $\lim(T \rightarrow \infty) (1/T)\Sigma_{t=1}^T \partial^2 E(\Phi_t(\theta^*))/\partial\theta\partial\theta'$. Gospodinov et al provide explicit expressions for the H and M matrices for the second HJD and Gospodinov et al(2012a) provide explicit expressions for the H and M matrices when using the first HJD. Both studies also provide explicit expressions for $\partial\Phi_t(\theta^*)/\partial\gamma$ and $\partial\Phi_t(\theta^*)/\partial\lambda$. Li et al(2010) provide corresponding expressions but use different notation. Under the null hypothesis that the model is correctly specified, then $\lambda^* = 0$ and $y_t(\gamma^*) = m_t(\theta^*)^+$ and both the H and M matrices simplify.

Gospodinov et al(2010) and Li et al(2010) derive the empirical test of a zero second HJD. Gospodinov et al show in Proposition 3 of their paper that under the null of a zero second HJD, Td^{+2} has a weighted $\chi^2(1)$ distribution. The $(N-KK)$ weights in the $\chi^2(1)$ distribution are the eigenvalues from:

$$A = P'U^{-1/2}SU^{-1/2}P \quad (9)$$

where P is (N, KK) orthonormal matrix where the columns are orthogonal to $U^{-1/2}D$, D is a (N, KK) matrix that equals $E[x_t(\partial y_t(\gamma)/\partial \gamma)']$, U is a (N, N) matrix equal to $E[x_t x_t']$, and S is a (N, N) matrix equal to $\Sigma_{j=-\infty}^{\infty} E[(x_t y_t(\gamma) - q_{t-1})(x_t y_t(\gamma) - q_{t-1})']$. All of the eigenvalues in the A matrix are positive. Gospodinov et al point out that the nonzero eigenvalues in the test of a zero second HJD proposed by Li et al(2010) are identical to the eigenvalues in A when the null of a zero second HJD is imposed. Gospodinov et al also note that the same approach is

used to test for a zero first HJD as in Jagannathan and Wang(1996) (see also Gospodinov et al(2012a)).

One of the problems of testing for a zero first HJD is that the simulation evidence in Ahn and Gadarowski(2004) suggests that the test tends to overreject models when N is large to T due to the use of the eigenvalues of the A matrix. As an alternative test of a zero first HJD, Gospodinov et al(2012a) propose the LM test of $\lambda=0_N$ which is given by:

$$T\lambda'U^{1/2}P(P'U^{-1/2}SU^{-1/2}P)^{-1}P'U^{1/2}\lambda \quad (10)$$

Under the null hypothesis of a zero first HJD, Theorem 1 of Gospodinov et al shows that the LM test has an asymptotic χ^2 distribution with N-KK degrees of freedom. Simulation evidence in Gospodinov et al shows that the LM test has good finite sample properties and is not subject to the overrejection problem of the Jagannathan and Wang(1996) test. I adapt the LM test to use as an additional model specification test within the second HJD framework.

Li et al(2010) and Gospodinov et al(2010) show that the model parameters γ and Lagrange Multipliers λ have an asymptotic normal distribution under the null of a misspecified model. Proposition 4 in Gospodinov et al shows that θ has an asymptotic normal distribution with covariance matrix given by $(1/T)\Sigma_\theta$ where $\Sigma_\theta = \sum_{j=-\infty}^{\infty} E(h_t h'_{t+j})$. The h_t series is given by $H^{-1}(\partial\Phi_t(\theta^*)/\partial\theta)$. Gospodinov et al point out that the asymptotic distribution of γ is normally distributed under the null hypothesis that the model is correctly specified but the asymptotic distribution of λ is no longer normally distributed. I use the asymptotic distribution of γ to examine hypotheses tests of the stochastic discount factor coefficients.

Li et al(2010) and Gospodinov et al(2010) develop pairwise model comparison tests of equation (7)¹³. Define model F as having K1 parameters to estimate and model G as having K2 parameters to estimate. When model F only contains a subset of factors in model G, then model F is a nested model of model G. When models F and G share some factors in common, then the two models are overlapping models. When models F and G share no factors in common, the two models are strictly non-nested. For the case of linear factor models, all of the models share the constant in common and so there are no strictly non-nested models.

Hansen et al(1995) show that the difference between d^{+2} for two models can be tested using a normal test. The difference between d^{+2}_F and d^{+2}_G has an asymptotic normal distribution with variance given by σ^2_d , where $\sigma^2_d = \sum_{j=-\infty}^{\infty} E(d_t d'_{t+j})$ and $d_t = [\Phi_t^F(\theta^*_F) - E(\Phi_t^F(\theta^*_F))] - [\Phi_t^G(\theta^*_G) - E(\Phi_t^G(\theta^*_G))]$ (see Gospodinov et al(2010)). Gospodinov et al point out that there are cases when we cannot use the normal test as $\sigma^2_d = 0$. The first case is when the two models have equal stochastic discount factor values ($y_t^F = y_t^G$) and the second case is when the two stochastic discount factor values are different from one another but both models are correctly specified and so $\Phi_t^F(\theta^*_F) = 0$ and $\Phi_t^G(\theta^*_G) = 0$.

For the nested model case, where I assume model F is a subset of model G, Gospodinov et al(2010) point out that $\sigma^2_d = 0$ under the null hypothesis that the two models have equal second HJD, which occurs only if the two models have equal stochastic discount factor values. As a result, for nested models one only needs to test the null hypothesis of equal stochastic discount factor values. The two models only have equal stochastic discount factor values when the slope coefficients on the (K2-K1) extra factors in model G are jointly

¹³ Kan and Robotti(2009) and Gospodinov et al(2012a) develop related model comparison tests using the first HJD and Kan et al(2012) develop model comparison tests based on the cross-sectional regression R^2 .

equal to zero (see Kan and Robotti(2009)). Li et al(2010) propose a weighted χ^2 test to test for equal second HJD for nested models. Gospodinov et al show that this test can be simplified by imposing the null hypothesis that $d^+_{\text{F}} = d^+_{\text{G}}$. Proposition 6 of their paper show that the hypothesis of zero slope coefficients on the (K2-K1) extra factors in model G can be tested by the $T(d^+_{\text{F}} - d^+_{\text{G}})$ statistic, which has a weighted $\chi^2(1)$ distribution with (K2-K1) weights. An alternative test is to use a Wald test which has an asymptotic χ^2 distribution with (K2-K1) degrees of freedom. In my study, I use the Wald test to examine if the two nested models have equal second HJD.

For the case where the two models are overlapping models $\sigma^2_{\text{d}} = 0$ whenever the two models have equal stochastic discount factor values or both models are correctly specified. Gospodinov et al(2010) recommend a sequential approach to test the equality of the second HJD for overlapping models. In the first step, a test of whether the two models have equal stochastic discount factor values is examined. Define K3 as the number of factors the two models F and G share in common. Kan and Robotti(2009) show that for two linear factor models, the stochastic discount factor values will be equal when the slope coefficients on the (K1-K3) factors in model F and the (K2-K3) factors in model G are jointly equal to zero. Gospodinov et al show that this hypothesis can be tested either by the statistic $T(d^+_{\text{F}} - d^+_{\text{G}})$, which has a weighted $\chi^2(1)$ distribution with (K1+K2-2K3) weights in Proposition 7 or by a Wald test which has an asymptotic χ^2 distribution with K1+K2-2K3 degrees of freedom.

The second step of the sequential approach is to test whether both models are correctly specified ($d^+_{\text{F}}=d^+_{\text{G}}=0$). Gospodinov et al(2010) show in Proposition 5 that under the null that both models are correctly specified then $T(d^+_{\text{F}} - d^+_{\text{G}})$ has a weighted $\chi^2(1)$ distribution with $2N-K1-K2$ weights. Gospodinov et al(2012a) in developing model comparison tests of the first HJD propose an alternative test based on an extension of the LM test in Lemma B2 (b) of their paper. This test examines if $\lambda_{\text{F}} = \lambda_{\text{G}} = 0_{\text{N}}$ and has asymptotic χ^2

distribution with $2N-K_1-K_2$ degrees of freedom. The LM test can be adapted to using the second HJD.

The final step in the sequential approach is to use the normal test. Gospodinov et al(2010) point out that to reject the null hypothesis of the equal second HJD between two models, we have to reject the tests in all three steps in the sequential approach at a specified significance level. In my study, I use the Wald test to examine whether the two overlapping models have equal stochastic discount factor values, use the LM test to examine if both models are correctly specified, and then use the normal test.

The pairwise model comparison tests only consider two models at a time. Gospodinov et al(2012a) develop multiple model comparison tests using the first HJD to evaluate models. The multiple model comparison tests of Gospodinov et al can be adapted to use for the second HJD. Gospodinov et al propose separate tests for multiple non-nested model comparison tests and multiple nested model comparison tests. Gospodinov et al develop a multiple non-nested model comparison test using the multivariate inequality tests of Wolak(1987,1989)¹⁴. Define $p+1$ as the number of models in the test where model 1 is the benchmark model and p is the set of alternative models. Model 1 has a squared second HJD given by d^{+2}_1 and d^{+2}_i is the squared second HJD of model i for the alternative $i=1,\dots,p$ models. The null hypothesis is that the benchmark model performs as well as the alternative models in terms of the squared second HJD and can be written as $\delta \leq 0_p$ where δ is a $(p,1)$ vector with individual elements equal to $\delta_i = d^{+2}_1 - d^{+2}_i$, and 0_p is a $(p,1)$ vector of zeros. The

¹⁴ Kan et al(2012) develop related multiple model comparison tests based on the cross-sectional regression R^2 . Chen and Ludvigson(2009) propose an alternative multiple model comparison test using the Hansen and Jagannathan distance measure following the bootstrap approach of White(2000) and Hansen(2005).

alternative hypothesis is that some alternative model has a lower second HJD than the benchmark model.

Assuming that δ has an asymptotic multivariate normal distribution models given by $N(0_p, \Omega_p)$, Gospodinov et al(2012a) show that a likelihood ratio (LR) test can be used for testing the null hypothesis. The LR test follows a “chi-bar-squared distribution” given by $\sum_{i=0}^p w_{p-i}(\Omega_p)X_i$ where the X_i are independent χ^2 variables with i degrees of freedom. The χ^2_0 is defined as 0 and the weights w_i sum to 1. Gospodinov et al show that the asymptotic p values of the LR test can be computed through numerical methods. The normality assumption of δ requires that $d_i^{+2} > 0$ and the stochastic discount factor values of each model are not equal to each other.

When comparing a benchmark model to a set of alternative models, Gospodinov et al(2012a) remove a number of the alternative models from the set. Any alternative models that are nested by the benchmark model are removed since $\delta_i \leq 0$ by construction. Any alternative model that is nested by another alternative model is also removed as the second HJD of the larger model will at least be as small as the smaller model. Any alternative model that nests the benchmark is removed since the normality assumption does not hold in this case under the null hypothesis that $\delta_i=0$. For the remaining alternative models that are left, the LR test is computed and the corresponding p value. In my study, I set each model to be the benchmark model and compute the corresponding LR test and p value.

For nested multiple model comparison tests, Gospodinov et al(2012a) point out that when the alternative models to the benchmark are nested with one another, then the smaller nested models can be excluded since the squared second HJD of the larger alternative model will be at least as small as the smaller models. In this case, the pairwise model comparison tests can be used. This is the case in all of the nested model comparison tests in my study. For example, when using the CAPM as the benchmark model, the FF, and Carhart models,

and the conditional CAPM and FF models are all nested within the conditional Carhart model and so the pairwise model comparison test between the CAPM and conditional Carhart model can be used for the multiple nested model comparison tests.

Gospodinov et al(2012a) suggest that for each model used as the benchmark, both the multiple non-nested model comparison tests and nested model comparison tests should be conducted separately. Gospodinov et al point out that if the two tests use a significance level of $\alpha/2$ and the null hypothesis cannot be rejected in either test, then the size of the joint test will be $\leq \alpha$ by the Bonferroni inequality.

B) Formation of the Primitive Assets

1) Size/DY Portfolios

I form the eighteen size/DY portfolios using the following approach at the start of each year between 1959 and 2010. All stocks on the LSPD are ranked by their market value at the start of the year and grouped into four portfolios. Within each size portfolio, I rank stocks on the basis of their DY at the start of the year. The DY for each stock is calculated as the sum of gross dividends in ex-div months during the past year divided by the price at the end of the year on LSPD. From July 1997 onwards, I use the net DY due to the abolition of the dividend tax credits by the U.K. government. The dividends and prices are corrected for capital changes using the capital change adjustment factors in LSPD. For companies with a non-zero DY, I form four portfolios where all portfolios contain an equal number of stocks as an approximation. I also form a separate portfolio of companies with a zero DY. For a company to be identified as a zero DY company, I require that companies have continuous monthly return observations during the year over which the DY is calculated. Due to the relatively small number of companies that have a zero DY in many of the years of my sample, I only form a zero DY portfolio for the smallest quartile of companies and form a second portfolio of zero DY companies that includes companies in the other three size

quartiles. I exclude companies with zero market values. I then calculate the monthly buy and hold returns during the next year for each portfolio. The initial weights in each portfolio are value weighted using the market value of the security at the start of the year.

I make a number of corrections and exclusions to the portfolio returns which I follow across forming the portfolios and factors. Where a security has missing return observations during the year, I 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. I correct for the delisting bias of Shumway(1997) by following the approach of Dimson et al(2003). A -100% return is assigned to the death event date on LSPD where the LSPD code indicates that the death is valueless. Prior to 1975, LSPD does not contain a complete return history of all companies. As a result, I only use the random 1/3 sample of companies in forming the portfolios up to 1975. The random 1/3 sample includes a random selection of 1/3 of companies that exist at the start of 1955 and 1/3 of companies that come onto the stock exchange during each of the subsequent years. I exclude investment trusts¹⁵ and from 1980 onwards, I also exclude foreign companies and secondary shares using data from the LSPD archive file.

2) Size/Beta Portfolios

I form the sixteen size/beta portfolios using the following approach at the start of each year between 1961 and 2010. All stocks on the LSPD are ranked by their market value at the start of the year and grouped into four portfolios. Within each size portfolio, I rank stocks on the basis of their beta relative to the value weighted market index and group into four portfolios. All portfolios contain an equal number of stocks as an approximation. The betas are estimated from the regression of the excess stock returns on a constant and the excess returns of the market index during the prior 60 months. I require companies to have

¹⁵ Investment trusts are equivalent to closed-end U.S. mutual funds.

continuous return data during the 60 prior months to estimate the betas. I exclude companies with zero market values. I then calculate the monthly buy and hold returns during the next year for each portfolio. The initial weights in each portfolio are value weighted using the market value of the security at the start of the year.

3) Size/BM Portfolios

I form the sixteen size/BM portfolios using the following approach at the start of each July between 1981 and 2010. All stocks on the LSPD are ranked by their market value at the end of June and grouped into four portfolios. Within each size portfolio, I rank stocks on the basis of their price to book (PB) ratio at the end of the previous year and group into four portfolios. All portfolios contain an equal number of stocks as an approximation. The PB ratio is collected from Datastream. I exclude companies with zero market values, zero and negative PB ratios, and financials. I then 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. I use the monthly returns of the size/BM portfolios between July 1981 and December 2010 in my empirical analysis.

The use of 18 size/DY portfolios, 16 size/beta portfolios, and 16 size/BM portfolios is a smaller number of portfolios than what is typically used in U.S. stock returns where 25 size/BM portfolios are frequently used. I form a smaller number of portfolios since it is only from January 1975 that LSPD contains complete return history of all U.K. companies. In addition, Dimson et al(2003) point out that the number of companies in the U.K. on the London Stock Exchange has reduced over time since 1955 in contrast to the increased number of U.S. companies. Due to the smaller number of companies available, I form a smaller number of portfolios for the size/DY, size/beta, and size/BM portfolios.

C) Formation of Factors in the Linear Factor Models

1) Factors in the Carhart model

I 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 1959 and 2010, I construct a value weighted portfolio of all stocks on LSPD by their market value at the start of the year. I calculate buy and hold monthly returns during the next year. I exclude companies with a zero market value.

I form the SMB and HML factors in the FF and Carhart models using the following approach. At the start of each year between 1959 and 2010, I rank all stocks on LSPD separately by their market value at the start of the year and by their DY at the start of the year. I next form two size groups (Small and Big) using the 70th percentile as the break point and three DY groups (Low, Medium, and High) using break points of the 40th and 60th percentiles. I then construct six portfolios of securities at the intersection of the size and DY groups (SL, SM, SH, BL, BM, BH). I calculate the monthly buy and hold return for the six portfolios during the next year. The initial weights are set equal to the market value weights at the start of the year. I exclude companies with a zero market value and a zero DY. The SMB factor is the difference in the average return of the three small firm portfolios (SL, SM, SH) and the average return of the three large firm portfolios (BL, BM, BH). The HML factor is the difference in the average return of the two high DY portfolios (SH and BH) and the average return of the two low DY portfolios (SL and BL).

I form the WML factor in the Carhart model using the following approach. At the start of each month between January 1959 and December 2010, all stocks on LSPD are ranked on the basis of their average return during months -12 to -2 . The top 1/3 (by average return) of companies is grouped into the Winners portfolio and bottom 1/3 of companies are grouped into the Losers portfolio. I calculate the average return on the Winners and Losers portfolios during the next month. I exclude companies with less than 12 past return

observations. The WML factor is the difference between the returns of the Winners and Losers portfolios.

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

I form the index-based models using a similar approach to Cremers et al(2012). I form the factors from two DY index portfolios, three size index portfolios, and six size/DY index portfolios. I form the index portfolios as follows. At the start of each year between 1959 and 2010, I rank all stocks on LSPD by their market value at the start of the year. I exclude stocks with the smallest 1%¹⁶ by market value when forming the index portfolios as the Russell indexes used by Cremers et al do not include the very smallest stocks.

I form the two DY index portfolios across all stocks. I rank all companies by their DY at the start of the year and group the top 1/3 (by DY) into a High/All portfolio and the bottom 1/3 into a Low/All portfolio. I exclude companies with a zero DY. I 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.

I form six size/DY index portfolios. For each size index, I rank all stocks in the index by their DY at the start of the year. I exclude companies with a zero DY. I group the top 1/3 of companies (by DY) into a High portfolio and the bottom 1/3 of companies into a Low portfolio. The six size/DY portfolios are Large/High, Large/Low, Mid/High, Mid/Low, Small/High, Small/Low.

For each index portfolio, I construct a value weighted portfolio and calculate monthly buy and hold monthly returns during the year. The initial weights are set to the market value

¹⁶ Dimson and Marsh(2001) refer to these stocks as micro-cap stocks.

weights at the start of the year. I 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, the difference in returns between the Small and Large size index portfolios (Small-Large), the difference in returns between the High/All and Low/All DY index portfolios (HML (All)), 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 (Small-Mid), the difference in returns between the Mid and Large size index portfolios (Mid-Large), the difference in returns between the Large/High and Large/Low size/DY index portfolios (HML (Large)), the difference in returns between the Mid/High and Mid/Low size/DY index portfolios (HML (Mid)), the difference in returns between the Small/High and Small/Low size/DY index portfolios (HML (Small)), and WML.

For the July 1981 and December 2010 subperiod, I also form the value/growth factors in the FF, Carhart, and the index-based models using the BM ratio. I follow a similar approach as above except the portfolios are formed at the start of July each year as in the size/BM portfolios. I exclude companies with negative and zero PB ratios and financials.

D) Information Variable

I use the lagged term spread as the information variable. The lagged term spread is the difference in the annualized yield of long term government bonds and the three-month Treasury Bill. I collect the long term bond yield from the U.K. country tables provided by the International Financial Statistics of the International Monetary Fund.

Table 1 Summary Statistics of Primitive Assets

Panel A Average Excess Returns					
	Zero	Low	2	3	High
Small	0.905	0.876	1.153	1.200	0.964
2		0.634	0.852	0.923	1.013
3		0.640	0.653	0.796	0.836
Big	0.355	0.383	0.475	0.688	0.735
Panel B Standard Deviations					
	Zero	Low	2	3	High
Small	6.007	5.125	4.729	4.877	6.233
2		4.861	4.532	4.874	5.523
3		5.282	4.933	5.301	5.906
Big	8.357	5.884	5.473	5.672	5.938

The table includes summary statistics of the monthly excess returns of 18 size/dividend yield (DY) portfolios between January 1959 and December 2010. The summary statistics include the mean (Panel A) and standard deviation (Panel B) of monthly excess returns (%). The portfolios of stocks are sorted by size in the rows (Small to Big) and DY in the column (Zero to High).

Table 2 Summary Statistics of Factors

Factors	Mean	Standard Deviation	Minimum	Maximum
Market	0.539	5.187	-26.005	46.856
SMB	0.185	3.113	-13.493	11.969
HML	0.276	2.382	-7.981	23.377
WML	0.632	3.099	-28.403	12.282
Large	0.501	5.289	-25.714	49.311
Small-Large	0.106	3.146	-11.443	12.114
HML (All)	0.338	3.774	-15.703	39.562
Small-Mid	-0.023	2.021	-13.190	8.176
Mid-Large	0.129	2.503	-9.573	10.628
HML (Large)	0.290	3.662	-12.121	16.385
HML (Mid)	0.230	4.016	-11.742	49.661
HML (Small)	0.349	2.806	-13.665	28.185

The table includes summary statistics of the monthly excess returns of factors used in the candidate stochastic discount factor models between January 1959 and December 2010. The summary statistics include the mean, standard deviation, minimum, and maximum values (%). 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. Small-Large, Small-Mid and Mid-Large 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. HML (All), HML (Large), HML (Mid), and HML (Small) are zero-cost portfolios of the value/growth effect in all companies, large companies, mid-cap companies, and small companies.

Table 3 Summary Statistics of Fitted Stochastic Discount Factor Models and Model Specification Tests

	$E(y)$	σ_y	$\text{Prop}(y < 0)$	d^+	$p(\lambda)$	$\text{SE}(d^+)$
CAPM	0.994	0.104	0.000	0.496 (0.000)	0.000	0.047
FF	0.994	0.182	0.320	0.475 (0.000)	0.000	0.045
Carhart	0.994	0.239	0.160	0.468 (0.000)	0.000	0.045
4-index	0.994	0.183	0.160	0.479 (0.000)	0.000	0.045
7-index	0.994	0.311	0.160	0.458 (0.000)	0.000	0.048
CAPM	0.994	0.292	0.160	0.472 (0.000)	0.000	0.050
FF	0.994	0.324	1.762	0.424 (0.000)	0.000	0.046
Carhart	0.994	0.445	2.884	0.402 (0.000)	0.000	0.046
4-index	0.994	0.394	1.282	0.418 (0.000)	0.000	0.046
7-index	0.994	0.494	2.564	0.392 (0.000)	0.000	0.049

The table reports the second Hansen and Jagannathan (1997) distance measure (HJD) and summary statistics of the fitted values of the candidate stochastic discount factors of linear factor models between January 1959 and December 2010. The conditional versions of the models are in bold. The N payoffs are the excess returns on 18 size/dividend yield (DY) portfolios, scaled excess returns of the size/DY portfolios, and the gross return on a three-month Treasury Bill. The model parameters in the linear factor models are estimated to minimize the second HJD. The summary statistics are the mean ($E(y)$) and standard deviation (σ_y) of the fitted stochastic discount factor values. The $\text{Prop}(y < 0)$ column is the proportion (%) of the fitted values of the stochastic discount factor that are negative. The d^+ column is the second HJD and the p value that tests the null hypothesis of a zero second HJD is in parentheses below. The $p(\lambda)$ column is the p value of the LM test, which examines if $\lambda = 0_N$. The $\text{SE}(d^+)$ column is the standard error of the second HJD under the null of a misspecified model. The test statistics are corrected for the effects of heteroskedasticity and serial correlation using the automatic lag selection (without prewhitening) method of Newey and West(1994).

Table 4 Model Comparison Tests

Panel A	FF	Carhart	4-index	7-index	CAPM	FF	Carhart	4-index	7-index
Pairwise									
CAPM	0.021 ¹	0.027 ¹	0.017	0.037 ²	0.024 ¹	0.067 ¹	0.085 ¹	0.072 ¹	0.093 ¹
FF		0.006	-0.003	0.016	0.004	0.046 ¹	0.064 ¹	0.051 ¹	0.072 ¹
Carhart			-0.010	0.009	-0.003	0.039 ¹	0.058 ¹	0.045 ¹	0.066 ¹
4-index				0.020	0.008	0.050 ¹	0.068 ¹	0.055 ¹	0.076 ¹
7-index					-0.012	0.030	0.048 ¹	0.035	0.056 ¹
CAPM						0.043 ¹	0.061 ¹	0.047 ¹	0.069 ¹
FF							0.018 ¹	0.005	0.026
Carhart								-0.013	0.008
4-index									0.021
Panel B			LR		<i>p</i> -value				
Multiple models									
CAPM			12.801						0.000
FF			11.727						0.001
Carhart			9.769						0.002
4-index			14.755						0.000
7-index			4.938						0.018
CAPM			7.379						0.005
FF			2.556						0.094
Carhart			0.343						0.424
4-index			3.861						0.054
7-index			0.000						0.605

¹ Significant at 5%

² Significant at 10%

The table reports the model comparison tests of Gospodinov et al(2010). The tests examine whether the squared second Hansen and Jagannathan(1997) distance measures (HJD) between two models are equal to each other between January 1959 and December 2010. The conditional versions of the models are in bold. The N payoffs are the excess returns on 18 size/dividend yield (DY) portfolios, scaled excess returns of the size/DY portfolios, and the gross return on a three-month Treasury Bill. The model parameters in the linear factor models are estimated to minimize the second HJD. Panel A of the table reports the difference in the squared second HJD between every pair of models in the pairwise model comparison tests. Where the difference is positive (negative), the model in the row has a higher (lower) second HJD than the model in the column. Panel B 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 lowest second HJD. The test statistics are corrected for the effects of heteroskedasticity and serial correlation using the automatic lag selection (without prewhitening) method of Newey and West(1994).

Table 5 Model Parameters and Lagrange Multipliers for a Subset of Models

Panel A Model Parameters								
FF	Constant	Market	SMB	HML				
γ_0 (γ_{0k})	1.027 (53.67) ¹	-2.693 (-2.27) ¹	-3.641 (-1.92) ²	-4.307 (-1.88) ²				
Carhart	Constant	Market	SMB	HML	WML			
γ_0 (γ_{0k})	1.305 (16.63) ¹	-3.002 (-1.83) ²	-0.277 (-0.11)	-12.409 (-3.27) ¹	-13.287 (-3.25) ¹			
γ_1 (γ_{1k})	-15.054 (-2.47) ¹	-41.965 (-0.71)	-295.212 (-3.15) ¹	176.612 (1.39)	244.277 (2.43) ¹			
Wald	0.000							
7-index	Constant	Large	Small-Mid	Mid-Large	HML (Large)	HML (Mid)	HML (Small)	WML
γ_0 (γ_{0k})	1.303 (14.66) ¹	-3.636 (-1.46)	-2.629 (-0.25)	1.774 (0.28)	-6.631 (-2.84) ¹	6.992 (1.18)	-11.755 (-2.05) ¹	-12.793 (-2.78) ¹
γ_1 (γ_{1k})	-15.294 (-2.21) ¹	-1.831 (-0.02)	-235.796 (-0.86)	-332.880 (-1.47)	157.159 (1.63)	-154.077 (-0.79)	137.89 4 (0.65)	214.870 (1.81) ²
Wald	0.000							
Panel B Lagrange Multipliers								
FF	FF	Carhart	Carhart	7-index	7-index			
Portfolio	λ	Portfolio	λ	Portfolio	λ			
Big/Low-scaled	-264.070 (-3.38) ¹	3/3-scaled	-543.967 (-2.70) ¹	2/Low-unscaled	-8.160 (-3.06) ¹			
Big/3-scaled	409.037 (3.42) ¹	2/High-scaled	502.970 (2.93) ¹	2/High-scaled	449.287 (3.08) ¹			
2/High-scaled	555.313 (3.55) ¹	Big/3-scaled	451.528 (4.32) ¹	Big/3-scaled	475.616 (4.98) ¹			
Big/High-scaled	-427.781 (-3.91) ¹	Big/High-scaled	-365.396 (-4.68) ¹	Big/High-scaled	-283.099 (-5.25) ¹			
T.Bill	-0.226 (-6.43)	T. Bill	-0.162 (-8.32)	T. Bill	-0.153 (-8.27)			

¹ Significant at 5%

² Significant at 10%

The table reports the model parameters (panel A) and the Lagrange Multipliers (panel B) for the unconditional FF model and the conditional Carhart and 7-index models between January 1959 and December 2010. The conditional versions of the models are in bold. The factors in the model are described in Table 2. The N payoffs are the excess returns on 18 size/dividend yield (DY) portfolios, scaled excess returns of the size/DY portfolios, and the gross return on a three-month Treasury Bill. The model parameters in the linear factor models are estimated to minimize the second HJD. Panel A of the table reports the coefficients in the stochastic discount factor (γ) and misspecification-robust t -statistics in parentheses. The γ_0 (γ_{0k}) row refers to the constant and slope coefficients on the factor excess returns. The γ_1 (γ_{1k}) row refers to the slope coefficients on the lagged term spread and the scaled factor excess returns. The Wald test examines the null hypothesis that the slope coefficients on the lagged information variable and scaled factor excess returns are jointly equal to zero. Panel B of the table reports the Lagrange Multipliers (λ) and t -statistics in parentheses of the five payoffs with the highest absolute t -statistics of λ . The Lagrange Multiplier of the gross Treasury Bill return is not normally distributed and so cannot be used to assess statistical significance. The Portfolio column in panel B refers to the corresponding payoff. The test statistics are corrected for the effects of heteroskedasticity and serial correlation using the automatic lag selection (without prewhitening) method of Newey and West(1994).

Table 6 Summary Statistics of Fitted Stochastic Discount Factor Models and Model Specification Tests: Size/Beta Portfolios

	$E(y)$	σ_y	$\text{Prop}(y<0)$	d^+	$p(\lambda)$	$\text{SE}(d^+)$
CAPM	0.994	0.099	0.000	0.440 (0.000)	0.001	0.058
FF	0.994	0.175	0.000	0.422 (0.000)	0.000	0.058
Carhart	0.994	0.411	1.833	0.385 (0.001)	0.000	0.056
4-index	0.994	0.404	1.167	0.389 (0.001)	0.000	0.055
7-index	0.994	0.454	2.000	0.377 (0.001)	0.000	0.050
CAPM	0.994	0.263	0.166	0.421 (0.001)	0.000	0.068
FF	0.994	0.296	1.166	0.369 (0.000)	0.001	0.058
Carhart	0.994	0.520	3.000	0.311 (0.018)	0.001	0.061
4-index	0.994	0.551	2.500	0.314 (0.021)	0.001	0.063
7-index	0.994	0.561	3.666	0.292 (0.011)	0.016	0.061

The table reports the second Hansen and Jagannathan (1997) distance measure (HJD) and summary statistics of the fitted values of the candidate stochastic discount factors of linear factor models between January 1961 and December 2010. The conditional versions of the models are in bold. The N payoffs are the excess returns on 16 size/beta portfolios, scaled excess returns of the size/beta portfolios, and the gross return on a three-month Treasury Bill. The model parameters in the linear factor models are estimated to minimize the second HJD. The summary statistics are the mean ($E(y)$) and standard deviation (σ_y) of the fitted stochastic discount factor values. The $\text{Prop}(y<0)$ column is the proportion (%) of the fitted values of the stochastic discount factor that are negative. The d^+ column is the second HJD and the p value that tests the null hypothesis of a zero second HJD is in parentheses below. The $p(\lambda)$ column is the p value of the LM test, which examines if $\lambda=0_N$. The $\text{SE}(d^+)$ column is the standard error of the second HJD under the null of a misspecified model. The test statistics are corrected for the effects of heteroskedasticity and serial correlation using the automatic lag selection (without prewhitening) method of Newey and West(1994).

Table 7 Summary Statistics of Fitted Stochastic Discount Factor Models and Model Specification Tests: Subperiod Results

	$E(y)$	σ_y	$\text{Prop}(y < 0)$	d^+	$p(\lambda)$	$\text{SE}(d^+)$
CAPM	0.994	0.100	0.000	0.516 (0.000)	0.011	0.088
FF	0.994	0.242	0.000	0.470 (0.003)	0.000	0.077
Carhart	0.994	0.297	0.000	0.463 (0.002)	0.002	0.078
4-index	0.994	0.135	0.000	0.508 (0.000)	0.004	0.086
7-index	0.994	0.425	2.825	0.415 (0.010)	0.000	0.072
CAPM	0.994	0.143	0.000	0.509 (0.000)	0.001	0.088
FF	0.994	0.313	0.847	0.429 (0.004)	0.003	0.072
Carhart	0.994	0.434	1.695	0.413 (0.005)	0.002	0.074
4-index	0.994	0.276	1.129	0.464 (0.000)	0.000	0.088
7-index	0.994	0.492	3.107	0.369 (0.006)	0.000	0.072

The table reports the second Hansen and Jagannathan (1997) distance measure (HJD) and summary statistics of the fitted values of the candidate stochastic discount factors of linear factor models between July 1981 and December 2010. The conditional versions of the models are in bold. The N payoffs are the excess returns on 16 size/book-to-market (BM) portfolios, scaled excess returns of the size/BM portfolios, and the gross return on a three-month Treasury Bill. The model parameters in the linear factor models are estimated to minimize the second HJD. The summary statistics are the mean ($E(y)$) and standard deviation (σ_y) of the fitted stochastic discount factor values. The $\text{Prop}(y < 0)$ column is the proportion (%) of the fitted values of the stochastic discount factor that are negative. The d^+ column is the second HJD and the p value that tests the null hypothesis of a zero second HJD is in parentheses below. The $p(\lambda)$ column is the p value of the LM test, which examines if $\lambda = 0_N$. The $\text{SE}(d^+)$ column is the standard error of the second HJD under the null of a misspecified model. The test statistics are corrected for the effects of heteroskedasticity and serial correlation using the automatic lag selection (without prewhitening) method of Newey and West(1994).

Acknowledgements

I am also grateful to C. Robotti, R. Kan, and N. Gospodinov for the use of their Matlab programs to implement some of the tests in the paper. Helpful comments received from an anonymous reviewer.

References

- Ahn, D.H., Cao, H.H. and S. Chretien, 2009, Portfolio performance measurement: A no arbitrage bounds approach, *European Financial Management*, 15, 298-339.
- Ahn,S.C., and C. Gadarowski, 2004, Small sample properties of the model specification test based on the Hansen-Jagannathan distance, *Journal of Empirical Finance*, 11, 109-132.
- Almeida, C. and R. Garcia, 2012, Assessing misspecified asset pricing models with empirical likelihood estimators, *Journal of Econometrics*, 170, 519-537.
- Alwathainani, A.M., 2012, Consistent winners and losers, *International Review of Economics and Finance*, 21, 210-220.
- 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.
- Chan, K.C., Karceski, J., and J. Lakonishok, 1998, The risk and return from factors, *Journal of Financial and Quantitative Analysis*, 33, 159-188.
- Chen, Z. and P.J. Knez, 1996, Portfolio performance measurement: Theory and applications, *Review of Financial Studies*, 9, 511-555.
- 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.
- 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).
- Cremers, M., Petajisto, A. and E. Zitzewitz, 2012, Should benchmark indices have alpha? Revisiting performance evaluation, *Critical Finance Review*, forthcoming.

Cuthbertson, K., Nitzsche, D. and N. O'Sullivan, 2008, Mutual fund performance: Skill or luck?, *Journal of Empirical Finance*, 15, 613-634.

Cuthbertson, K., Nitzsche, D. and N. O'Sullivan, 2010, Mutual fund performance: Measurement and evidence, *Financial Markets, Institutions, and Instruments*, 19, 95-187.

Davis, J.R., Fletcher, J. and A. Marshall, 2012, Testing index-based models in U.K. stock returns, *Working Paper*, University of Strathclyde.

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

Dimson, E., Nagel, S. and G. Quigley, 2003, Capturing the value premium in the U.K. 1955-2001, *Financial Analysts Journal*, 59, 35-45.

Dybvig, P. and S.A. Ross, 1985, Differential information and performance measurement using a security market line, *Journal of Finance*, 40, 383-399.

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.

Farnsworth, H., Ferson, W.E., Jackson, D. and S. Todd, 2002, Performance evaluation with stochastic discount factors, *Journal of Business*, 75, 473-504.

Ferson, W.E., 2003, Tests of multifactor pricing models volatility bounds and portfolio performance, In Constantinides GM, Harris M, Stultz R(eds) *Handbook of the Economics of Finance*, Elsevier Science Publishers, North Holland, 743-800.

Ferson, W.E., 2012, Investment performance: A review and synthesis, In Constantinides, G.M., Harris, M., and R. Stultz (eds) *Handbook of the Economics of Finance*, Elsevier Science Publishers, North Holland, forthcoming.

Fletcher, J., 1994, The mean-variance efficiency of benchmark portfolios: UK evidence, *Journal of Banking and Finance*, 18, 673-685.

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

Fletcher, J. and J. Hillier, 2002, An examination of the economic significance of stock return predictability in U.K. stock returns, *International Review of Economics and Finance*, 11, 373-392.

Glosten, L. and R. Jagannathan, 1994, A contingent claims framework for analyzing the performance of portfolio managers, *Journal of Empirical Finance*, 1, 133-160.

Gospodinov, N., Kan, R. and C. Robotti, 2010, On the Hansen-Jagannathan distance with a no-arbitrage constraint, *Working Paper*, Federal Reserve Bank of Atlanta.

Gospodinov, N., Kan, R. and C. Robotti, 2012a, Chi-squared tests for evaluation and comparison of asset pricing models, *Working Paper*, Concordia University.

Gospodinov, N., Kan, R. and C. Robotti, 2012b, Further results on the limiting distribution of GMM sample moment conditions, *Journal of Business and Economic Statistics*, forthcoming.

Hansen, P.R., 2005, A test for superior predictive ability, *Journal of Business and Economic Statistics*, 23, 365-380.

Hansen, L.P., Heaton, J. and E.G. Luttmer, 1995, Econometric evaluation of asset pricing models, *Review of Financial Studies*, 8, 238-274.

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

Hansen, L.P. and S.F.R. Richard, 1987. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587-613.

Harrison, M and D. Kreps, 1979. Martingales and arbitrage in multi-period securities markets, *Journal of Economic Theory* 20, 381-408.

- Hodrick, R.J. and X. Zhang, 2001. Evaluating the specification errors of asset pricing models, *Journal of Financial Economics* 62, 327-376.
- Jagannathan, R. and Z. Wang, 1996. The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3-53.
- Kan, R. and C. Robotti, 2008, Specification tests of asset pricing models using excess returns, *Journal of Empirical Finance*, 15, 816-838.
- Kan, R. and C. Robotti, 2009, Model comparison using the Hansen-Jagannathan distance, *Review of Financial Studies*, 22, 3449-3490.
- Kan, R., Robotti, C. and J. Shanken, 2012, Pricing model performance and the two-pass cross-sectional regression methodology, *Journal of Finance*, forthcoming.
- 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.
- Lewellen, J.W., Nagel, S. and J. Shanken, 2010, A sceptical appraisal of asset-pricing tests, *Journal of Financial Economics*, 96, 175-194.
- 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*, forthcoming.
- Ludvigson, S. and S. Ng, 2007, The empirical risk-return tradeoff: A factor analysis approach, *Journal of Financial Economics*, 83, 171-222.
- Merton, R.C., 1981, On market timing and investment performance. I. An equilibrium theory of value for market forecasts, *Journal of Business*, 54, 363-406.

- Nagel, S. and K. Singleton, 2011, Estimation and evaluation of conditional asset pricing models, *Journal of Finance*, 66, 873-909.
- Newey, W. K. and K.D. West, 1994, Automatic lag selection in covariance matrix estimation, *Review of Economic Studies*, 61, 631-653.
- Ross, S.A., 1978, A simple approach to the valuation of risky streams, *Journal of Business* 51, 153-475.
- 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*, forthcoming.
- White, H., 2000, A reality check for data snooping, *Econometrica*, 68, 1097-1126.
- 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.