US Efficient Factors in a Bayesian Model Scan Framework

Key words: Bayesian analysis, Model Scan, Factor Models

Abstract

Ehsani and Linnainmaa (2021) show that time-series efficient investment factors in US stock returns span and earn 40% higher Sharpe ratios than the original factors. We examine the impact of these efficient factors on factor model comparison tests in U.S returns. Using the Bayesian model scan approach of Chib, Zeng and Zhao(2020), and Chib, Zhao and Zhou(2022) on a set of 26 factors across the period 1972-2022 we show that that the optimal asset pricing model is an 8-factor model which contains efficient versions of the Market factor, Value factor (HML) and long-horizon behavioural factor (FIN). Our findings show that these efficient factors enhance the performance of U.S factor model performance. The top performing asset pricing model does not change in recent data.

1. Introduction

Asset pricing models can be improved by either adding factors that expand the efficient frontier or by enhancing the mean-variance efficiency of the existing factors. Ehsani and Linnainmaa (2022) follow the second approach by examining if the autocorrelation present in factor returns can be used boost the Sharpe ratio of an investment factor. The authors follow the framework of Ferson and Siegel (2001) to create so-called efficient factors in which factor weightings are conditioned on the information contained in lagged returns. Using this framework Ehsani and Linnainmaa (2022) show that time-series efficient factors significantly outperform the standard factors in U.S stock returns. In the Fama and French (2015) five-factor model, all efficient factors earn higher Sharpe ratios than the original factors and largely span the standard factors.

Classical tests of asset pricing models examine portfolio efficiency by comparing squared Sharpe ratios (Sharpe, 1992). The tests of Gibbons, Ross and Shanken (1989) compare the maximum squared Sharpe ratio, S^2 (r,f), of a portfolio formed from test assets r and factors f to that of the portfolio of factors, S^2 (f). Barillas and Shanken (2017) outline that in this framework, when comparing two models' factors, the model with the higher S^2 (f) produces smaller pricing errors for any test assets. The authors outline how test assets tell us nothing about model comparison in these cases, beyond what we learn by examining the extent to which each model prices the factors in the other model. The logic of the Barillas and Shanken (2017) framework allow tests of model comparison to focus on the maximization of investment factors' squared Sharpe ratios in order to identify the optimal asset pricing model (O'Connell, 2022). The most recent tests of model comparison in the U.S stock returns when the extent of model mispricing is gauged by the squared Sharpe ratio improvement measure is that of Barillas, Kan, Robotti and Shanken (2020).

Ehsani and Linnainmaa (2022) outline how their result is not specific to the Fama and French (2015) five-factor model: time-series efficiency improves Sharpe ratios of the factors in all popular asset pricing models. The objective of this paper is to examine the optimal combination of factors in the top performing asset pricing models in US stock returns over the period 1972-2022 when these efficient factors are considered.

Given the large number of factors we are aiming to examine we utilise the Bayesian approach for model comparison tests developed by Barillas and Shanken (2018). The Bayesian method is useful as we have a large number of factors to examine, as a list of predetermined models may exclude important combinations. This approach evaluates factor models on the basis of their Marginal Likelihoods (ML), from which the posterior probabilities of the models can be calculated. Chib, Zeng and Zhao (2020) provide a critique of the Barillas and Shanken (2018) approach and show that their method of calculating ML is not appropriate. Chib et al propose an alternative approach to calculating ML that can be used for relative model comparison tests. The attraction of their approach is that the ML can be solved analytically as relies on multivariate normality. Chib, Zhao and Zhou (2022) use this approach to run a model scan for comparing models in a set of 12 U.S factor returns.

We use the model scan approach building on Chib et al., (2020), and Chib et al., (2022) to examine the performance of different models that can be formed from 13 investment factors and their efficient counterparts over the sample period is between July 1972 and December 2022. We use the first 10% of the sample as a training sample as in Chib et al.,(2022) to estimate the hyperparameters for the prior distribution, and we use the remaining 90% to conduct the model comparison tests. We also compare Student-t distributed factor models with Gaussian distributed factor models over our period of analysis.

There main finding of our study is we find that the optimal asset pricing model is an 8-factor model which contains efficient versions of the Market factor, Value factor (HML) and long-horizon behavioural factor (FIN) along with the following original factors {Market, BAB, MGMT, PERF, PEAD}. This shows that the efficient factor transformation of Ehsani and Linnainmaa (2022) does have an impact on our model comparison tests. This transformation enhances existing factor models in U.S stock returns. When we examine optimal change points the model identified changes very slightly in recent data.

There are three main contributions of our study. First, we are the only authors to examine if the efficient factors developed by Ehsani and Linnainmaa (2022) have an impact on model comparison tests in U.S stock returns. Second, we complement the Bayesian model scan studies of Barillas and Shanken (2018), Chib et al (2021), and Chib et al (2022) in U.S. stock returns by conducting a model scan on an extended group of factors. Third, as well as conducting a model scan across the whole sample period, we also consider whether the best factor model in more recent data is different from the best factor model using all data using the approach of Chib, Zhao and Zhou (2021). We find that the optimal change point in the sample period is December 1997. There is a small change in the best factor model in recent data.

The paper is organized as follows. Section 2 presents the research method and describes the data used in my study. Section 3 reports the empirical results. The final section concludes.

2. Research Method

2.1 Model Comparison Framework

Ross (1978), Harrison and Kreps (1979), and Hansen and Richard (1987) show that if the Law of One Price (LOP) holds in financial markets, then a stochastic discount factor (m_{t+1}) exists such that:

$$\mathrm{E}(m_{t+1}X_{it+1}|Z_t) = p_{it} \qquad \text{for i=1,….N} \tag{1}$$

where X_{it+1} is the payoff of asset i at time t+1, p_{it} is the cost of asset i at time t, Z_t is the information set used by investors, and N is the number of primitive assets. If $m_{t+1} > 0$, then financial markets also satisfy the No Arbitrage (NA) opportunities in financial markets (Cochrane (2005))[1]. If the asset payoffs are excess returns then equation (1) implies that:

$$E(m_{t+1}r_{it+1}) = 0 \qquad \text{for } i=1,\ldots,N \tag{2}$$

where r_{it+1} is the excess return of asset i at time t+1.

Most asset pricing models specify a candidate model for the stochastic discount factor (y_{t+1}). The most popular models are linear factor models, where the candidate stochastic discount factor is given by:

$$y_{t+1} = \alpha + \sum_{k=1}^{K} b_k f_{k+1} \tag{3}$$

where α and b_k are the constant and slope coefficients in the stochastic discount factor, f_{k+1} are the values of the factors at time t+1, and K is the number of factors in the model. The slope coefficient (b_k) tells us whether factor k is important in pricing the primitive assets given the other factors in the model (Cochrane (2005)). These models are linear, as they define the securities returns to be a linear combination of factor returns weighted by the securities factor exposures. In this study we only consider factor models with traded factors where factors included are constructed from market trading or accounting data.

When we evaluate linear factor models using equation (2), we face a difficulty in identifying all the coefficients of the stochastic discount factor in equation (3). As a result of this, we must select a specific normalization for the stochastic discount factor. We choose the normalization followed by Chib and Zeng (2020) so that the expected value of the stochastic discount factor is set equal to 1, and equation (3) becomes:

$$y_{t+1} = 1 + \sum_{k=1}^{K} b_k fd_{kt+1} \tag{4}$$

where fd_{kt+1} is the demeaned value of factor k at time t+1.

Dybvig and Ingersoll (1982) and Ferson and Jagannathan (1996) demonstrate that linear factor models establish an analogous relationship between expected returns and betas. Cochrane (2005) and Ferson (2019) establish that stochastic discount factors, expected return-to-beta ratios, and mean-variance frontiers represent equivalent frameworks. Cochrane shows that if the linear factor model in equation (4) satisfies the pricing restrictions in equation (2), then:

$$E(r_{it+1}) = \sum_{k=1}^{K} \beta_{ik}\lambda_k \tag{5}$$

[1] The stochastic discount factor will only be unique if markets are complete.

where β_{ik} is the factor beta of asset i relative to factor k, and λ_k is the factor risk premium of factor k. Define b is a (K,1) vector of stochastic discount factor coefficients (b_k), λ is a (K,1) vector of factor premiums (λ_K), and V_f is the (K,K) covariance matrix of the factors, Cochrane shows that:

$$b = -V_f^{-1} \lambda \tag{6}$$

This mispricing framework is challenged by Barillas and Shanken (2017) who show that for relative model comparison tests for traded linear factor models, the choice of test assets is irrelevant for a number of metrics. Any linear factor model should be able to correctly price the test assets, and any excluded factors from the model. When the union of all the factors in each model is included in the investment universe, the test assets drop out of the analysis and are, therefore, irrelevant for model comparison. To illustrate, consider two models A and B with factors f_A, and f_B, and a set of test asset excess returns r. Using the maximum squared Sharpe (1966) (Sh^2) ratio to compare models, then for model A the metric is $Sh^2(f_A, f_B, r) - Sh^2(f_A)$, and for model B $Sh^2(f_A, f_B, r) - Sh^2(f_B)$. Given the $Sh^2(f_A, f_B, r)$ is fixed across models, then we can compare the relative performance of the models using the Sh2(f_A) and Sh2(f_B) measures alone. Chib and Zeng (2020) extend this test asset irrelevance framework to the stochastic discount factor approach in relative model comparison tests.

Since equations (2) and (5) hold for linear factor models in the form of equation (4), this implies that:

$$E(F_{t+1}) = \lambda \text{ for factors included in the model,} \tag{7}$$

$$\text{and } E(F^*_{t+1}) = \beta\lambda \text{ for factors excluded from the model}$$

where β is a (L,K) matrix of factor betas, F_{t+1} is the K factors included in the model, and F^*_{t+1}, represents the L factors excluded from the model. The restrictions of equation (7) can be cast into a regression framework in which the excess returns of the factors incorporated in the model are regressed against a constant term, while the excess returns of the factors not included in the model are regressed against the excess returns of the included factors, with the removal of the intercept term. The exclusion of the intercept[2] in the second regression imposes the zero pricing error restriction on the excluded factors.

Barillas and Shanken (2018) derive a Bayesian model comparison test on the basis of Marginal Likelihoods (ML) that can be used to compare the performance of a large number of traded factor models simultaneously. Chib, Zeng, and Zhao (2020) present a critique of the Barillas and Shanken approach, focusing on their selection of prior distributions on model-specific nuisance parameters in the regression framework. In response, Chib et al. (2020) introduces a more comprehensive framework that can be utilized within the constrained regression

[2] If the intercept was included it would capture the Jensen (1968) alpha of the excluded factors. The Jensen alpha is identical to the stochastic discount factor alpha in this setup as $E(y_{t+1}) = 1$. See Ferson(2019).

framework outlined in equation (7). Chib and Zeng (2020) begin their approach by initially considering a model that encompasses all K+L factors within the set of factor models under comparison. They introduce a prior distribution for this comprehensive model that incorporates all factors. Subsequently, they derive the prior distribution for any model that involves a subset of these factors. The log ML of a candidate model (m_j) is given by:

$$\text{Log } ML(m_j) = \log ML(F) + \log ML(F^*) \tag{8}$$

Assuming that the factor data follows a multivariate normal distribution, Chib et al.,(2022) show that the log ML of each model can be solved analytically in Proposition 5 of their paper. Chib, Zhao and Zhou (2022) apply the results of Chib et al.,(2020) to evaluate model comparison tests of all models that can comprise the factor models of Fama and French (2018), Hou et al (2015), Stambaugh and Yuan (2017), and Daniel et al (2020), which they identify as winner factor models given their performance in historical tests of model comparison. There are J = $2^{K+L} - 1$ potential factor models that can be constructed and is defined as a model scan. The models can be compared using their posterior probabilities, assuming that each factor model has an equal prior probability of 1/J as:

$$\text{Posterior Probability}_j = ML_j \,/\, \Sigma_{j=1}^{J} ML_j \tag{9}$$

where ML_j is the marginal likelihood of model j. The approach of Chib et al.,(2020) uses a training sample to estimate the hyperparameters for the prior distributions. We use the first 10% of our sample period as the training sample. Chib et al.,(2022) also derive the posterior distribution of the various parameters in the regression framework, from which we can then derive the posterior distribution of the stochastic discount factor coefficients in equation (6). To assess the capability of a specific factor model in pricing an omitted factor, we can refer to the two regressions, omitting the time t+1:

$$F_1^* = \alpha_1 + \beta_1^U F_K + e_1^U \tag{9a}$$

$$F_1^* = \beta_1^R F_K + e_1^R \tag{9b}$$

where F_1^* is the excess return of the excluded factor l, β_1^U and β_1^R are (1,K) vectors of the betas from the unrestricted and restricted regressions, α_l is the Jensen(1968) performance measure, and e_1^U, and e_1^R are the residuals from the two regressions. Chib et al., (2020) estimate the log ML for each equation. According to Jeffrey's rule, if the difference in the log ML exceeds 1.15, the null hypothesis of a zero alpha can be dismissed.

A recent paper by Chib, Zhao, and Zhou in 2021 expands the scope of model comparison tests to address the possibility that the most suitable factor model for recent data might differ from the one identified using the entire dataset. Chib et al.,(2022) posit that the observed divergence might result from significant shifts caused by the widespread growth of the internet, the concept of adaptive efficient markets presented by Lo (2004), or the impact of the publication effect as outlined by McLean and Pontiff (2016). To tackle this issue, Chib et al.,(2022) introduce the concept of employing the model scan approach in conjunction ML estimation to determine the optimal change point within the dataset. The overall sample period is split into two subperiods for a given change point. Chib et al.,(2021) use December 1996, June 1997, December 1997, June 1998, December 1998, June 1999, and December 1999 due to the internet revolution. Define t^* as a given split point. The model scan is run on both subperiods, and the ML is calculated for each potential model as ML_{1j} and ML_{2j}. Chib et al.,(2022) then calculate the ML for a given split point t^* as:

$$ML_{t*} = (1/J^2)\, \Sigma^{J}_{j=1}\Sigma^{J}_{j=1}\, ML_{1J}\, ML_{2J} \tag{10}$$

The ML_{t*} is calculated for each split point, and the optimal split point is given by the highest ML_{t*}. Given the optimal split point, the model scan is then run on the second subperiod to identify the best factor model in the most recent data.

2.2 Efficient Factors

As outlined by Ferson and Siegel (2001), conditioning information is present when the optimal solution may be a function of information received about the probability distribution of future outcomes. We must consider that even though factor models can be conditionally efficient, the factors themselves may be unconditionally mean-variance inefficient. When the factors are not unconditionally MVE, the asset pricing model may not explain a fairly priced asset's mean return with constant coefficients. To solve this problem, one could increase the number of factors in the model in an attempt to minimise the non-zero alpha. Instead, Ehsani and Linnainmaa (2022) construct unconditionally MVE factors to conduct a valid test of the model by the means of time-series regressions.

The authors assume that a factor's return follows an AR(1) process to form unconditionally MVE factors. In other words, the prior month return of a given factors may contain information useful in deciding the optimal weight on that factor in the following month. The investor needs just three parameters the factor's unconditional mean, variance, and autocorrelation and the factor's prior return to generate a factor's time-series efficient version. A time-series efficient factor exploits the autocorrelation in factor returns; it times the original factor to minimize variance while maintaining the expected return. Time-series efficient factors may deliver higher Sharpe ratios relative to the original factors and contain all the information found in the original factors.

Ehsani and Linnainmaa (2022) use the framework of Ferson and Siegal (2001) to construct time series efficient factors. Time series efficient factors being a portfolio of factors where the weights of each factor in the portfolio are a function of conditioning information, which in our cases is the prior month return of the factor. Starting from a single risky asset with a return of

$$\tilde{R} = \mu(\tilde{S}) + \tilde{\varepsilon}, \tag{11}$$

where $\tilde{R}$ is the risky asset's return in excess of the risk-free rate, $\tilde{S}$ is the predictor (signal), $\mu(\tilde{S})$ is the expected excess return conditional on the signal, and $\tilde{\varepsilon}$ is the random noise net of the signal with a mean of zero and a variance of $\sigma^2_{\varepsilon}(\tilde{S})$. The efficient strategy invests $x(\tilde{S})$ in the risky asset and the remainder, $1 - x(\tilde{S})$, in the risk-free asset. The unconditional expected excess return and variance of this investment strategy are given by

$$\mu_p = E\,[x(\tilde{S})\,\mu(\tilde{S})], \tag{12}$$

$$\sigma^2_p = E\,[x^2(\tilde{S})\,(\mu^2(\tilde{S}) + \sigma^2_{\varepsilon}(\tilde{S}))] - \mu^2_p \tag{13}$$

Ferson and Siegel (2001) show that the portfolio that minimizes σ^2_p for a given conditional expectation μ_p invests $x(\tilde{S})$ in the risky asset,

$$x(\tilde{S}) = \frac{\mu_P}{\partial}\frac{\mu(\tilde{S})}{\mu^2(\tilde{S}) + \sigma^2_{\varepsilon}(\tilde{S})} \tag{14}$$

Here μ_P denotes the unconditional expected factor returns obtained from the original factor. The conditional expected portfolio returns $\mu(\tilde{S})$, assuming an AR(1) model is used to condition the time-series Efficient factor on, and the constant ζ are defined below.

$$\partial = \frac{\mu^2(\tilde{S})}{\mu^2(\tilde{S}) + \sigma^2_{\varepsilon}(\tilde{S})} \tag{15}$$

This weighting program produces a unique mean-variance efficient portfolio. That being no other portfolio has the same unconditional return at a lower unconditional variance (Ferson and Siegel, 2001).

Ehsani and Linnainmaa (2022) focus on time-series efficiency, using information embedded in the factors' realized returns. This case gives a closed-form solution for the MVE transformation and for the expected efficiency gain or increase in Sharpe ratio. The new factors the authors construct, using information only in factors' past returns, are weak-form efficient in the sense of Fama (1970). Ehsani and Linnainmaa (2022) assume that past returns are related to future returns but unrelated to variance. Specifically, we assume that returns follow a homoscedastic autoregressive process,

$$\tilde{R}_t = \mu + \rho\, \tilde{R}_{t-1} + \varepsilon_t \tag{16}$$

$$\mathrm{var}[\varepsilon_t \mid R_{t-1}] = \sigma_\varepsilon^2 \tag{17}$$

The factor's conditional expected return under this model is $\mu(\tilde{S}) = \mu + \rho\, \tilde{R}_{t-1}$. Using equations (13) and (15), the investor's optimal weight on the factor is

$$x(S_t) = \mu_p \frac{SR^2+1}{SR^2+\rho^2} \frac{\mu_p(1-\rho)+ \rho r_{t-1}}{\left(\mu_p(1-\rho)+\rho r_{t-1}\right)^2+\sigma_\varepsilon^2} \tag{18}$$

In this equation, μ_p is the factor's unconditional mean, SR is the unconditional Sharpe ratio, ρ is the autocorrelation coefficient, and $\sigma^2_\varepsilon = (1 - p^2)\, \sigma^2$ is the constant variance of the noise term. We define time-series efficient factor as the portfolio that invests $x(S_t)$ (from equation (5)) on the original factor. A time-series efficient HML, for example, would be the return on a portfolio that optimally times HML given, in this derivation, its month $t-1$ return. In our empirical work, like Ehsani and Linnainmaa (2022), we use month $t - 1$ return as our conditioning information. The optimal weight on a given factor depends on the factor's mean, standard deviation, and first-order autocorrelation.

2.3 Data

Our focus in this study is on U.S. factors The factor data used in this study comes from the Ken French website. We obtain the updated value factor (HML_M), quality minus junk factor (QMJ) along with the Betting Against Beta (BAB) factor from the AQR database. We obtain the behavioural factors from Lin Suns' homepage. All factors are denominated in USD. The approach for constructing the factor portfolios follows Fama and French (1993, 2012). The market factor (MKT) consists of value-weighted returns of all available (and valid) securities on the U.S market less the risk-free rate.

The factors in the FF6 model are the excess returns on the market index, and zero-cost portfolios of the size (SMB), value (HML), profitability (RMW), investment (CMA), and momentum (MOM) effects in stock returns. We include the Betting against Beta (BAB) factor of Frazzini and Pedersen(2014) along with the two mispricing factors termed Management (MMGT), and Performance (PERF) of Stambaugh and Yuan(2017). The short-horizon behavioural factor (PEAD) and long-horizon behavioural factor (FIN) of Daniel, Hirshleifer, and Sun (2020) are also included.

Table 1 reports summary statistics of the excess factor returns between July 1972 and December 2022. The summary statistics include the average excess return (%), standard deviation (Std Dev), and the *t*-statistic of the null hypothesis that the average excess factor returns are equal to zero.

Table 1 Summary Statistics of Factors

	Mean	Std Dev	*t*-statistic
Market	0.590	4.612	3.146[1]
SMB	0.164	3.005	1.344
HML	0.333	3.101	2.644[1]
RMW	0.301	2.323	3.190[1]
CMA	0.328	2.031	3.974[1]
Mom	0.617	4.360	3.481[1]
BAB	0.858	3.430	6.150[1]
HMLm	0.343	3.703	2.278[1]
QMJ	0.405	2.357	4.224[1]
MGMT	0.643	2.708	5.840[1]
PERF	0.584	3.879	3.706[1]
PEAD	0.576	1.912	7.410[1]
FIN	0.734	3.899	4.629[1]

The table reports summary statistics of factors between July 1972 and December 2022. The summary statistics include the average excess returns (%) and standard deviation (Std Dev) of the factors. The t-statistic column is the t-statistic of the null hypothesis that the average excess factor returns are equal to zero.
1 Significant at 5%
2 Significant at 10%

Table 1 shows that all of the factors have significant positive average excess returns, except for the SMB factor which has an insignificant positive return. The BAB and FIN factors have the largest average excess returns at 0.858%, and 0.734% respectively. The mispricing factors of Stambaugh and Yuan (2017) have significant positive average excess returns. All but three of the factors in our set have a *t*-statistic higher than 3, which is the cut-off *t*-statistic recommended by Harvey, Liu and Zhu (2016) to control for multiple testing.

3. Empirical Results

Barillas, Kan, Robotti and Shanken (2020) develop asymptotically valid tests of model comparison when the extent of model mispricing is gauged by the squared Sharpe ratio improvement. Using this framework the authors conduct the most recent tests of model comparison in the U.S stock returns from 1972 to 2015. In these tests they find that a variant of the Fama and French (2018) six-factor model, with a monthly-updated version of the usual value spread, emerges as the dominant model. We conduct these tests on the same subset of factor models as Barillas et al., (2020) from the period 1972 to 2022. Table 1A contained in Appendix A1 presents the results of these tests. Naturally, our results are almost identical to those of Barillas et al., (2020). We find that the variant of the Fama and French (2018) six-factor model, with a monthly-updated version of the value factor emerges as the dominant model. The only difference in our study is that we find the second highest performing model to be the Chib, Zeng, Zhao (2020) five-factor model instead of the Hou et al. (2015) (HXZ) q factor model. The Fama and French (2018) six factor model remains the third highest performing model followed by the Fama and French five factor model and Stambaugh and Yuan (2017) (SY) four factor model. From Panel B we can see that the differences in Sharpe ratios of the top performing models is not statistically significant. Full details on these asymptotically valid tests of model comparison can be found in Barillas et al., (2020)3.

Turning to a Bayesian method of model comparison we begin by running the model scan using all 13 original factors. There are 8,192 possible models, and we assign an equal prior probability to them all as in Chib and Zeng (2020), and Chib et al.,(2022). Table 2 reports the empirical results. Panel A of the Table reports the results for the top 6 models in terms of the highest posterior probability. Panel A includes the posterior probability of each model, the ratio of the posterior probability to the prior probability, and the difference in log marginal likelihoods (ML) between the best model (M1) to that of another model. Chib et al., (2022) outline that if the difference in log ML ≤ 1.15 according to the Jeffrey's rule, then the best model is indistinguishable from the alternative model. Panel B reports the identity of the factors in the top 6 models from the model scan.

Table 2 Model Scan of 13 Factors

Panel A: **Top Models** Model	Posterior Probability	Posterior/Prior	ML
1	0.1514	1240.12	
2	0.14667	1201.4	0.03172
3	0.05889	482.368	0.94425
4	0.05718	468.325	0.9738
5	0.03536	289.627	1.45437
6	0.0329	269.501	1.52639

Panel B: **Factors**								
1	Market	BAB	MGMT	PERF	PEAD	FIN		
2	Market	SMB	BAB	MGMT	PERF	PEAD	FIN	
3	Market	CMA	BAB	MGMT	PERF	PEAD	FIN	
4	Market	RMW	CMA	BAB	MGMT	PERF	PEAD	
5	Market	BAB	HMLm	MGMT	PERF	PEAD	FIN	
6	Market	SMB	CMA	BAB	MGMT	PERF	PEAD	FIN

The table reports the results of the Bayesian model scan of 13 factors in U.S. stock returns. The sample period is July 1972 and December 2022. The first 10% of the sample period is used for the training sample, and the model scan is then conducted on the remaining 90% of the sample period. Panel A reports the posterior probability, the ratio of posterior probability to prior probability, for the top 6 models. The ML column is the difference in the log

3 We are thankful to Professor Cesare Robotti for the provision of the code online to run these model comparison tests using Sharpe ratios.

ML of the best model and the next best models 2 to 7. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan.

Table 2 shows that the best factor model in the model scan is a six-factor model with a posterior probability of 0.1514. The next best model has a posterior probability of 0.014667. The next four best models have a posterior probability that ranges between 0.05889 and 0.05718. The ratio of the posterior probability to prior probability shows a substantial increase for the six best models. The differences in log ML in Panel A of Table 2 are all below 1.15 for the next top 3 performing models and so the best model is statistically indistinguishable from the other top 3 models. The results in panel A of Table 2 are very similar to that from Chib et al(2022).

Panel B of Table 2 shows that the best model includes the Market, BAB, MGMT, PERF, PEAD, and FIN factors. The next top models contain the majority of factors outlined with the addition of either the SMB or CMA factor. The role of the market index in the best factor models is consistent with Harvey and Liu(2021). The SMB factor is included in 2 of the top models, which is surprising given the low average excess returns of the SMB factor in Table 1.

Table 2 suggests that the best model in terms of posterior probability is a six-factor model. Chib et al(2022) derive the posterior distribution of the factor premiums in a given factor model. We use 10,000 simulation draws for generating the posterior distribution of the factor premiums, and the corresponding stochastic discount factor coefficients in the best factor model.

Table 3 reports the summary statistics of the posterior distribution of the factor premiums (Panel A), and stochastic discount factor coefficients (Panel B). The summary statistics include the mean, standard deviation (Std Dev), median, and 2.5% and 97.5% percentiles of the posterior distribution.

Table 3 Summary Statistics of the Posterior Distribution of the Best Model Risk Factors

Panel A: **Premiums**	Mean	Std Dev	Median	2.50%	97.50%
Market	0.668	0.198	0.665	0.282	1.058
BAB	0.880	0.153	0.882	0.578	1.178
MGMT	0.578	0.116	0.578	0.350	0.811
PERF	0.640	0.168	0.642	0.309	0.969
FIN	0.540	0.083	0.539	0.380	0.705
PEAD	0.678	0.170	0.679	0.344	1.011

Panel B: **SDF Coeffs**	Mean	Std Dev	Median	2.50%	97.50%
Market	-7.494	1.286	-7.477	-10.067	-5.007
BAB	-6.225	1.651	-6.214	-9.507	-3.009
MGMT	-8.478	1.865	-8.453	-12.246	-4.903
PERF	-5.167	1.268	-5.152	-7.690	-2.715
FIN	-17.501	2.978	-17.488	-23.465	-11.731
PEAD	-5.803	1.599	-5.783	-8.986	-2.750

The table reports the summary statistics of the posterior distribution of the factors in the best model from the Bayesian model scan of 13 factors in U.S. stock returns. The sample period is July 1972 and December 2022. The first 10% of the sample period i used for the training sample, and the model scan is then conducted on the remaining 90% of the sample period. Panel A reports the summary statistics of the posterior distribution of the factor premiums (%), and panel B reports the summary statistics of the posterior distribution of the stochastic discount factor coefficients. The summary statistics include the mean, standard deviation (Std Dev), median, 2.5% and 97.5% percentiles using 10,000 simulation draws.

Panel A of Table 3 shows that the BAB factor has the largest mean factor premium at 0.880%, followed by the PEAD factor at 0.678%. All of the factor premiums are significantly positive using the 95% percentile interval with the exception of the Market factor. In Panel B of Table 3 all of the mean stochastic discount factor coefficients are negative for each factor, and significantly negative using the 95% percentile intervals. This finding suggests that all six factors play an important role in the stochastic discount factor in pricing assets given the other factors in the model (Cochrane (2005)). The Market factor plays an important role even where the mean factor premium is not significantly positive.

Similar to Qiao, Wang, and Lam (2022) we will now compare the results from our model scan where the joint distribution of the factor data is assumed to be Gaussian to results where this distribution is assumed to be multivariate t. Qiao et al (2022) find strong evidence that the Student-t distributed global factor pricing models significantly outperform the Gaussian distributed ones highlighting the importance of using multivariate Student-t distributions to model the fat tails in global risk factor data. We rerun our model scan of our set of 13 U.S factors assuming multivariate t with three degrees of freedom. In this case when the joint distribution of our risk factors follows a Student-t distribution, to calculate the marginal likelihood of each contending model, we first use an initial portion of our data as the training sample to get prior distribution of the parameters of the factor model. The Markov chain Monte

Carlo (MCMC) method is then used to get the posterior distribution of the parameters and calculate their posterior means, with which we further calculate the marginal likelihood of the factor model4. For full details on the calculation of the marginal likelihood under the multivariate-t assumption see Chib et al,. (2020).

Table 4 Model Scan of 13 Factors assuming Multivariate-t Factor Distribution

Panel A: **Top Models** Model	Posterior Probability	Posterior/Prior	ML
1	0.4071	3335.17	
2	0.1446	1185.32	1.03451
3	0.1086	889.67	1.32143
4	0.0477	391.412	2.14252
5	0.0429	352.209	2.24805
6	0.0391	320.686	2.34182

Panel B: **Factors**								
1	Market	BAB	MGMT	PERF	PEAD	FIN		
2	Market	SMB	BAB	MGMT	PERF	PEAD	FIN	
3	Market	MOM	BAB	MGMT	PERF	PEAD	FIN	
4	Market	BAB	HMLm	MGMT	PERF	PEAD	FIN	
5	Market	CMA	BAB	MGMT	PERF	PEAD	FIN	
6	Market	SMB	MOM	BAB	MGMT	PERF	PEAD	FIN

The table reports the results of the Bayesian model scan of 13 factors in U.S. stock returns when the joint distribution of factors is assumed to follow a multivariate t distribution. The sample period is July 1972 and December 2022. Panel A reports the posterior probability, the ratio of posterior probability to prior probability, for the top 6 models. The ML column is the difference in the log ML of the best model and the next best models 2 to 7. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan.

Similar to Qiao, Wang, and Lam (2022) we find that the multivariate t assumption identifies the same top performing models as when we assuming a gaussian joint distribution of factors. We also find increased support for this top performing six-factor model as our posterior probability increases to 0.4071 from 0.1514. The differences in log ML in Panel A of Table 4 are below 1.15 for only the second highest performing model. When we assumed a Gaussian distribution on our factors we found that the top four models were statistically indistinguishable from each other . This again shows increased support for our top two performing models when we assuming fat tails on our U.S factor data.

4 We are thankful to Professor Siddhartha Chib for providing the code online for which we used to run the model scan assuming multivariate normality for the joint distribution of factors.

We now rerun our model scan including our original factors along with their efficient counterparts calculated using the framework of Ehsani and Linnainmaa (2022). This gives us a starting collection of 26 factors. Assuming a multivariate normal distribution on the joint factor data there are now 67,108,864 possible models, and we assign an equal prior probability to them all as in Chib and Zeng(2020), and Chib et al(2022). Given the fact that Ehsani and Linnainmaa (2022) show the efficient factor transformation allows for a significant increase in the Sharpe ratio provided by U.S factors we would expect these factors to improve the Sharpe ratio provided by a given model and therefore be present in an optimal asset pricing model.

Table 5 reports the empirical results. Panel A of the Table reports the results for the top 6 models in terms of the highest posterior probability. Panel A includes the posterior probability of each model and the difference in log marginal likelihoods (ML) between the best model (M1) to that of another model. Panel B reports the identity of the factors in the top 6 models from the model scan.

Table 5 Model Scan of 26 Factors

Panel A: **Models**	Posterior Probability	ML
Model		
1	0.161	
2	0.057	1.043
3	0.054	1.087
4	0.047	1.238
5	0.031	1.642
6	0.028	1.752

Panel B: **Factors**									
1	Market	BAB	MGMT	PERF	PEAD	Market_{ef}	HMLm_{ef}	FIN_{ef}	
2	Market	RMW	BAB	MGMT	PEAD	Market_{ef}	HMLm_{ef}	FIN_{ef}	
3	Market	BAB	MGMT	PEAD	FIN	Market_{ef}	HMLm_{ef}	PEAD_{ef}	FIN_{ef}
4	Market	BAB	MGMT	PERF	PEAD	Market_{ef}	HMLm_{ef}	MGMT_{ef}	FIN_{ef}
5	Market	BAB	MGMT	PERF	Market_{ef}	MOMef	HMLm_{ef}	PEAD_{ef}	FIN_{ef}
6	Market	MGMT	PERF	PEAD	Market_{ef}	BAB_{ef}	HMLm_{ef}	PEAD_{ef}	FIN_{ef}

The table reports the results of the Bayesian model scan of 26 factors in U.S. stock returns. This set is made up of 13 factors and their efficient counterparts. The sample period is July 1972 and December 2022. The first 10% of the sample period is used for the training sample, and the model scan is then conducted on the remaining 90% of the sample period. Panel A reports the posterior probability. The ML column is the difference in the log ML of the best model and the next best models 2 to 6. Panel B presents the identity of the factors in the top 6 models from the Bayesian model scan.

Table 5 shows that the best factor model in the model scan is an eight-factor model with a posterior probability of 0.161 indicating that there is large support from the data for this particular model from our possible set. The next six best models have a posterior probability that ranges between 0.057 and 0.028. The ratio of the posterior probability to prior probability shows a substantial increase for the seven best models. The difference in log ML in panel A of Table 2 are above 1.15 for models 4-6, however it is below 1.15 for the top three models and so the best model is statistically indistinguishable from the other top two models.

We can see from Panel B that the efficient factors are included in the top performing models. More specifically the efficient versions of the Market factor, Value factor (HML) and long-horizon behavioural factor (FIN) are included in the top performing model along with the following original factors {Market, BAB, MGMT, PERF, PEAD}. All of the original factors have been retained in the top performing model however when the FIN factor undergoes the efficient factor transformation it improves the performance of the asset pricing model to the extent that the original factor is no longer required. The inclusion of the efficient value factor ($HMLm_{ef}$) is not surprising given that Ehsani and Linnainmaa (2022) note the large increase in Sharpe performance for this factor when its weight is conditioned on its previous returns.

The analysis so far has run the model scan using all the available data. Chib et al(2021) argues that the choice of the relevant factor model might change for a number of reasons for example technological change, or the adaptive efficient market hypothesis (Lo(2004)) among others. We adapt the Bayesian approach of Chib et al(2021) to identify the optimal split point among the set of split points used in their study. Table 5 reports the log ML for each of the split points (t*) for December 1996, June 1997, December 1997, June 1998, December 1998, June 1999, and December 1999.

Table 6 Test of Optimal Change Point

t*	Log ML
Dec-96	11482.36
Jun-97	11492
Dec-97	11493.82
Jun-98	11547.33
Dec-98	11524.47
Jun-99	11447.32
Dec-99	11455.66

The table reports the results of the Bayesian tests of the optimal split sample point following the approach of Chib et al(2021). The overall sample period is July 1983 and June 2021. The first 10% of the sample period is used for the training sample, and the model scan is then conducted over the two subperiods where sample split points are set to December 1996, June 1997, December 1997, June 1998, December 1998, June 1999, and December 1999. The table reports the log of the marginal likelihoods (ML) for each sample split period.

Table 6 shows that the optimal sample split point is June 1998 as it has the highest marginal likelihood. This result is consistent with Chib et al(2021). Chib et al point out that this split period occurs during the internet revolution, and the boom in tech stocks. Given the optimal split point of June 1998, we then repeat the model scan tests during the June 1998 and December 2022 sample period. Table 6 reports the top 6 models with the highest posterior probabilities (panel A), and the identity of the factors (panel B) in the top models.

Table 7 Model Scan of 26 Factors in Recent Data

Panel A: **Models**	Posterior Probability	ML
Model		
1	0.019	
2	0.017	0.080
3	0.010	0.583
4	0.010	0.630
5	0.006	1.123
6	0.006	1.175

Panel B: **Factors**									
1	Market	BAB	PERF	PEAD	$Market_{ef}$	$HMLm_{ef}$	$MGMT_{ef}$	$PEAD_{ef}$	FIN_{ef}
2	Market	BAB	MGMT	PERF	PEAD	$Market_{ef}$	$HMLm_{ef}$	$PEAD_{ef}$	FIN_{ef}
3	Market	BAB	PEAD	FIN	$HMLm_{ef}$	$PEAD_{ef}$	FIN_{ef}		
4	Market	BAB	MGMT	PERF	PEAD	$Market_{ef}$	$HMLm_{ef}$	$MGMT_{ef}$	FIN_{ef}
5	Market	BAB	PERF	PEAD	$Market_{ef}$	MOM_{ef}	$HMLm_{ef}$	$MGMT_{ef}$	$PEAD_{ef}$
6	Market	MGMT	PERF	$Market_{ef}$	$HMLm_{ef}$	$MGMT_{ef}$	$PEAD_{ef}$	FIN_{ef}	

The table reports the results of the Bayesian model scan of 26 factors in U.S. stock returns. This set is made up of 13 factors and their efficient counterparts. The sample period is June 1998 and December 2022

Table 7 shows that the best factor models using the most recent data are similar to that for the entire sample. The best factor model in for the whole sample is a lot more dominant than the factor model in Table 6 in terms of a much higher posterior probability. The best model has the highest posterior probability of 0.01862. There is then drop in the posterior probabilities of the next best factor models. There is no statistical difference in the performance of the best factor model and the 4 next best models. We see the same factors emerge in the optimal model as we did in the full sample. In this subperiod the efficient version of the behavioural factor (FIN) is included in more of the top models.

4. Conclusions

We use the Bayesian model scan approach of Chib et al (2020), and Chib et al (2022) to examine model comparison tests among a set of U.S. factors and their efficient counterparts. Our objective was to examine if the efficient factor transformation of Ehsani and Linnainmaa (2022) had an impact on model comparison tests in U.S stock returns.

First, similar to Chib and Zend (2020), we find that the best factor model during the whole sample period with the highest posterior probability is a six-factor model, which includes the Market, BAB, MGMT, PERF, FIN, and PEAD factors. The posterior probability is large at 0.1514, however the performance of the best model is statistically indistinguishable from the next best three factor models in terms of posterior probability. All six factors in the best model play a significant role in the stochastic discount factor given the other factors in the model.

Similar to Qiao, Wang, and Lam (2022) when we assume that the joint distribution of our risk factors follows a Student-t distribution, we identify the same six factor model as the optimal model with increased support from our data.

Second, we find that the best factor model from the Bayesian model scan when the efficient versions of the original factors are included in the starting set does include some of these efficient factors. We find that the efficient version of the Market factor, Value factor (HML) and long-horizon behavioural factor (FIN) are included in the optimal asset pricing model over our sample period model along with the following original factors {Market, BAB, MGMT, PERF, PEAD}. This finding indicates that the efficient factor transformation of Ehsani and Linnainmaa (2022) has an impact on model comparison tests in U.S stock returns.

Third, we find that the best factor model does not change with the use of the most recent data. We find that the optimal split point in the sample is June 1998. The best factor model in the June 1998 and December 2022 period is a nine-factor model, which consists of the factors found in the whole sample scan plus the efficient version of the behavioural factor (FIN).

Our study suggests that the efficient factor transformation should be performed and included in model comparison tests in the asset pricing literature. Our analysis has used the Bayesian model scan under the multivariate normal distribution of Chib et al(2020), and Chib et al(2021a). Further work could be extended to include the efficient factors in a classical asset pricing framework.

References

Barillas, F., Kan, R., Robotti, C. and J. Shanken, 2020, Model comparison with Sharpe ratios, *Journal of Financial and Quantitative Analysis*, 55, 1840-1874.

Barillas, F. and J. Shanken, 2017, Which alpha?, *Review of Financial Studies*, 30,1316-1388.

Barillas, F. and J. Shanken, 2018, Comparing asset pricing models, *Journal of Finance*, 73, 715-754.

Chib, S., Lin, Y.C., Pukthuanthong, K. and X. Zeng, 2021, Slope factors outperform: Evidence from a large comparative study, *Working Paper*, Washington University in St Louis.

Chib, S. and X. Zeng, 2020, Which factors are risk factors in asset pricing? A model scan approach, *Journal of Business and Economic Statistics*, 38, 771-783.

Chib, S., Zeng, X. and L. Zhao, 2020, On comparing asset pricing models, *Journal of Finance*, 75, 551-577.

Chib, S., L. Zhao and G. Zhou, 2021, Finding risk-factors …. without using old data, *Working Paper*, Washington University in St Louis.

Chib, S., Zhao, L. and G. Zhou, 2022, Winners from winners: A tale of risk factors, *Working Paper*, Washington University in St Louis.

Cochrane, J.H., 2005, *Asset pricing: Revised edition*, Princeton NJ: Princeton University Press.

Daniel, K.D., Hirshleifer, D. and L. Sun, 2020, Short- and long-horizon behavioural factors, *Review of Financial Studies*, 33, 1673-1736.

Dybvig, P. and J.E. Ingersoll, 1982, Mean-variance theory in complete markets, 55, 233-251.

Ehansi, S. and J.T. Linnainmaa, 2021, Time-series efficient factors, *Working Paper*, Tuck School of Business.

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, 2015, A five-factor asset pricing model, *Journal of Financial Economics*, 116, 1-22.

Fama, E.F. and K.R. French, 2018, Choosing factors, *Journal of Financial Economics*, 128, 234-252.

Ferson, W.E., 2019, *Empirical asset pricing: models and methods*, MIT Press.

Ferson, W.E. and R. Jagannathan, 1996, Econometric evaluation of asset pricing models,

Ferson, Wayne E., and Andrew F. Siegel, 2001, The efficient use of conditioning information in portfolios, Journal of Finance 56, 967–982.

Frazzini, A. and L.H. Pedersen, 2014, Betting against beta, *Journal of Financial Economics*, 111, 1-25.

Gibbons, M. R.; S. A. Ross; and J. Shanken. "A Test of the Efficiency of a Given Portfolio." Econometrica, 57 (1989), 1121–1152.

Hansen, L.P. and R. Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy*, 99, 225-262.

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.

Harvey, C.R., 2017, Presidential address: The scientific outlook in financial economics, *Journal of Finance*, 72, 1399-1440.

Harvey, C.R., 2020, Replication in financial economics, *Critical Finance Review*, 8, 1-9.

Harvey, C.R. and Y, Liu, 2021, Lucky factors, *Journal of Financial Economics*, forthcoming.

Harvey, C.R., Liu, Y. and H. Zhu, 2016, … and the cross-section of expected returns, *Review of Financial Studies*, 29, 5-68.

Hou, K., Xue, C. and L. Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies*, 28, 650-705.

Hou, X. Xue, C. and L. Zhang, 2020, Replicating anomalies, *Review of Financial Studies*, 33, 2019-2133.

Jensen, M.C., 1968, The performance of mutual funds in the period 1945-1964, *Journal of Finance* 23, 389-416.

Kan, R., Robotti, C. and J. Shanken, 2013, Pricing model performance and the two-pass cross-sectional regression methodology, *Journal of Finance*, 68, 2617-2013.

Lo, A.W., 2004, The adaptive markets hypothesis, *Journal of Portfolio Management*, 30, 15-29.

McLean, R.D. and J. Pontiff, 2016, Does academic publication destroy stock return predictability, *Journal of Finance*, 71, 5-32.

O'Connell, M. (2022), "Model comparison in German stock returns", *Journal of Economic Studies*, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/JES-05-2022-0261

Pontiff, J. and A. Woodgate, 2008, Shares issuance and cross-sectional returns, *Journal of Finance*, 43, 921-945.

Ross, S.A., 1978, A simple approach to the valuation of risky streams, *Journal of Business*, 51, 153-475.

Sharpe, W.F., 1966, Mutual fund performance, *Journal of Business* 39, 119-138.

Stambaugh, R.F. and Y. Yuan, 2017, Mispricing factors, *Review of Financial Studies*, 30, 1270-1315.

Appendix A1

Table 1A. USA Tests of Equality of Squared Sharpe Ratios

Panel A: Difference in Squared Sharpe Performance

Model	Carhart	FrazPed	FF5	HXZ	SY	FF6	CZZ	AsFraz
FF3	-0.042	-0.051	-0.059	-0.059	-0.066	-0.087	-0.089	-0.095
Carhart		-0.009	-0.017	-0.017	-0.024	-0.046	-0.047	-0.053
FrazPed			-0.008	-0.008	-0.015	-0.036	-0.038	-0.043
FF5				0	-0.007	-0.028	-0.03	-0.035
HXZCP					-0.007	-0.028	-0.03	-0.035
SY						-0.022	-0.023	-0.029
FF6							-0.002	-0.007
CZZ								-0.005

Panel B: p-Values

Model	Carhart	FrazPed	FF5	HXZ	SY	FF6	CZZ	AsFraz
FF3	0	0.033	0	0.001	0.015	0	0.001	0.001
Carhart		0.723	0.5	0.48	0.482	0	0.01	0.008
FrazPed			0.747	0.744	0.691	0.185	0.165	0.113
FF5				0.975	0.844	0.11	0.086	0.06
HXZCP					0.84	0.117	0.095	0.065
SY						0.567	0.535	0.453
FF6							0.013	0.342
CZZ								0.066

Panel A shows the differences between the (bias-adjusted) sample squared Sharpe ratios (column model - row model) for various pairs of models. The models are presented from left to right and top to bottom in order of increasing squared Sharpe ratios. The diagonal elements are the sample squared Sharpe ratio differences between the model in that column and the next-best model. Model references are the following: Fama and French (1993)(FF3), Carhart(1997)(Carhart), Frazzini and Pedersen (2014)(FrazPed), Fama and French(2015)(FF5), Hou et al. (2015) (HXZ), Stambaugh and Yuan (2017) (SY), Fama and French (2017) (FF6) ,Chib, Zeng, Zhao (2020) (CZZ), Asness and Frazzini (2013) (ASFraz).