Structural Breaks, Model Uncertainty and Factor Selection

Siddhartha Chib[a,*], Simon C. Smith[b]

[a] *Olin School of Business, Washington University in St. Louis*
[b] *Federal Reserve Board*

Abstract

This paper addresses the non-standard problem of detecting multiple structural breaks when the data-generating model in each regime is uncertain, with an application to factor selection in empirical asset pricing. Detection is based on the marginal likelihood of break points, obtained by a novel integration over all possible pairings of models across regimes, from all possible models within regimes. The optimal break points maximize this marginal likelihood. Applying this method to the six Fama-French factors on monthly data from 1963–2023, the analysis identifies three breaks—1982, 1998, and 2009—and a shift toward more parsimonious models after 1998. Before 1998, five or six factors are selected, but two afterward. Thus, with breaks, there is a move to parsimony, which has implications for the factor zoo literature. Moreover, within each regime, all omitted factors are spanned by the ones selected. Incorporating breaks also leads to substantially different weight allocations in the maximum Sharpe ratio risk factor portfolio.

Keywords: Model comparison, Factor models, Structural breaks, Bayesian analysis.
JEL classifications: G12, C11, C12, C52, C58

1. Introduction

'*US small-cap stocks are suffering their worst run of performance relative to large companies in more than 20 years [...] The Russell 2000 index has risen 24% since the beginning of 2020, lagging the S&P 500's more than 60% gain over the same period. The gap in performance upends a long-term historical norm in which fast-growing small-caps have tended to deliver punchier returns for investors who can stomach the higher volatility.*"[1] *(Financial Times, 2024)*

The empirical literature on asset pricing has proposed a large number of factors that claim to explain the cross-section of expected stock returns (Cochrane, 2011). More recently, the field has been dealing with how to handle this proliferation of factors. Various potential solutions have been offered (Feng *et al.*, 2020).

[⋆] January 2025

[*] Corresponding author

Email addresses: chib@wustl.edu (Siddhartha Chib), simon.c.smith@frb.gov (Simon C. Smith)

[1] This quote is from a March 27, 2024 Financial Times article entitled US small-caps suffer worst run against larger stocks in more than 20 years.'

This paper presents an intuitively simple point of view that has somehow been overlooked in the literature. If the set of factors that explain the cross section of expected returns is varying over time, it is critical to account for this feature when evaluating which factors are relevant at any given time. There is little doubt that the set of risk factors change. This can be seen with the Fama-French set of risk factors, which have changed twice over time. Different economic explanations can support such changes, for example, the publication effect of Schwert (2003), and/or the adaptive efficient market hypothesis of Lo (2004). The set of risk factors may also change due to, for example, new technologies, shifts in monetary policy regimes, or regulations. Using all available historical data will tend to pick up factors that were important at some point in the past, but are not risk factors at present. As a simple example, imagine that only two factors are relevant for the first half of the sample and that two different factors are relevant in the second half. The common approach in the literature of using all the historical data will tend to suggest that all four factors are relevant for the entire sample, when in fact no more than two are relevant at any given time. This may partly explain the problem of the "factor zoo" (Harvey *et al.*, 2016; Hou *et al.*, 2020), as well as the declining performance of risk factors in a comprehensive set of anomalies (McLean and Pontiff, 2016). Therefore, it is important to consider the possibility of time variation when selecting factors.

If one knew the time at which the set of factors changes, one could discard the old irrelevant data with a subsample split. In reality, however, this date is not known and therefore must be estimated. Green *et al.* (2017), for example, impose a predetermined subsample split in the early 2000s and find that the number of relevant characteristics has declined over time. Furthermore, the longer the sample period under consideration, the more likely it is that there may be multiple times at which the set changes, which further complicates the problem. This setting is technically challenging because one needs to estimate both the times at which the set of relevant factors changes and the set of relevant factors within each subperiod. In other words, both the asset pricing model and the parameters of that model change. This setting is more complex than standard breakpoint problems in which the model parameters shift after a break but the model itself (i.e. the selected factors) remains unchanged. A widely applied approach for this setting was developed in Chib (1998), first applied in the finance setting by Pástor and Stambaugh (2001) and subsequently in many other papers. Standard breakpoint problems have been applied to a range of issues in empirical asset pricing, such as return predictability (Viceira, 1997; Lettau and Van Nieuwerburgh, 2008; Rapach *et al.*, 2010; Smith and Timmermann, 2021), estimating time-varying risk premia (Pástor and Stambaugh, 2001; Smith and Timmermann, 2022), and dating the integration of world equity markets (Bekaert *et al.*, 2002). In this paper, we propose a solution to this challenging problem by devising the first method (Bayesian or frequentist) that can simultaneously estimate both the times at which the model changes and how the parameters of the model change, taking the guesswork out of how to determine the subsample splits (or regimes).

In our methodology, change point detection recognizes model uncertainty. Following Chib and Zeng (2020), the uncertainty about risk factors is dealt with by considering every possible split of the given set of factors into risk factors $\boldsymbol{x}$ and non-risk factors $\boldsymbol{y}$. For a given set of K factors, there are $J = 2^K - 1$ such splits. Each such split defines a particular asset pricing model with its own set of parameters. Now in the context of change points, this model uncertainty is multiplied. We do not know the risk factors in any of the segments induced by the change points. Consider the case of one break. To address this uncertainty about the risk factors, we consider the joint distribution of the data given a particular change point t_1, models $\mathbb{M}_{j_1,1}$ on the left with parameters $\boldsymbol{\theta}_{j_1,1}$, and $\mathbb{M}_{j_2,2}$ on the right with parameters $\boldsymbol{\theta}_{j_2,2}$ for $(j_1, j_2) = 1, ..., J = 2^K - 1$. Then, using the priors on the parameters from Chib and Zeng (2020), which have been extensively vetted, we find the marginal likelihood of the entire data given $(t_1, \mathbb{M}_{j_1,1}, \mathbb{M}_{j_2,2})$. We then marginalize out the models by considering all possible pairings of models (j_1, j_2). This gives the marginal likelihood of the entire data given just that change-point, an idea introduced in the regression context in Chib (2024). We then repeat this calculation for a grid for possible change points and multiply by the prior on those change points (taken to be discrete uniform) to get the posterior distribution of change points up to a normalizing constant. From this posterior distribution, we take the optimal change point to be the one with the modal posterior probability. Given the optimal change points we get the risk factors in each of the segments by the model scanning approach of Chib and Zeng (2020) applied to the data in each segment. The same approach extends in a straightforward way to two or more change points. Later in the paper, we discuss details about this approach including the computational intensity and possibilities for dealing with the large K case.

In our empirical analysis, we focus on the six-factor model of Fama and French (2018). The model scan is therefore over 63 models, including the popular risk-factor collections such as the 3- and 5-factor Fama-French models, but it also includes all other combinations of risk-factors that have not previously been considered. Using monthly data from July 1963 through December 2023, our method identifies three breaks corresponding to a regime lasting 15 years on average. The breaks occur in 1982, 1998, and 2009. These break dates correspond to the end of the "monetarist policy experiment" implemented by Paul Volcker between 1979 and 1982, the Internet revolution and the tech boom on the NASDAQ (Griffin *et al.*, 2011),[2], and the Global Financial Crisis (GFC).

The set of risk factors changes after each of these breaks. At least five factors are selected in the first two regimes (up to 1998), while only two factors are selected since. In the current (post-2009) regime, only the market and profitability factors are selected. In contrast, the preferred model when using all historical data is a four-factor model that excludes size and value, which shows that failing to discard pre-break data can lead to a risk factor set being selected that is not the relevant one for pricing in the current regime. This selected model is unable to price one of the omitted factors – size – using the whole sample of available data, highlighting its shortcomings. Furthermore, using the entire data sample, our approach reveals that the momentum factor is not priced by the Fama-French 5-factor model; and the momentum, investment, and profitability factors are not priced by the Fama-French 3-factor model.

While dense factor models are most informative until 1998, we find evidence of a clear shift toward more parsimonious models since 1998. This parallels the finding of Kelly *et al.* (2019) who use Instrumented Principal Components Analysis to document that just five latent factors can outperform existing factor models.

In every regime, each of the omitted factors is priced by the selected factors, suggesting that they are spanned by the smaller subset of selected factors and can therefore be confidently excluded. The maximum Sharpe ratio portfolio shows clear evidence of time-varying weight allocations across the different factors, a feature that is concealed when precluding breaks. In addition, our methodology would be useful for detecting any change in the current set of risk factors in the future.

Finally, our methodology provides regime-specific estimates of factor risk premia and their price of risk. A small subset of studies that estimate time-varying risk premia include Ferson and Harvey (1991); Freyberger *et al.* (2020), Gu *et al.* (2020), Gagliardini *et al.* (2016), Ang and Kristensen (2012), and Adrian *et al.* (2015). Mounting empirical evidence of sizeable risk premia associated with these factors has important implications for investment strategies and has markedly changed the investment landscape, leading to the proliferation of mutual funds specializing in certain investment styles such as small caps or value stocks. The appeal of such strategies depends not only on the magnitude of the associated risk premia, but also on the stability of their risk premia over time. Factor premia may time-vary due to investors differing in sophistication or investment objectives, enabling the marginal investor to differ across stocks and over time for a given stock. Individual investors can form mean-variance

[2]This break also coincides with a period of dramatic changes in market efficiency that has been documented by Chordia *et al.* (2011).

portfolios, while others may pursue very large payoffs. Some investors may follow "buy-and-hold" strategies, and others may periodically rebalance to target certain weights. We find clear time-variation in the risk premia for all six factors since 1963. For example, the value premium and size premium have declined over time and neither factor is selected in the current regime (Fama and French, 2021). The implied weights on the value factor in the maximum Sharpe ratio portfolio therefore declined over time, indicating that high allocations to value stocks have become notably less attractive over time.

Bessembinder *et al.* (2021) estimate factor risk premia using a fixed 60-month rolling window and document clear time-variation in the number of factors selected over time. However, as we show in our empirical analysis, a rolling window leads to factors entering and exiting the SDF very frequently, sometimes on a monthly basis. The economic motivation for this behavior, however, is difficult to justify. This is why a formal method is needed to identify the set of risk factors that is stable within a regime, but is allowed to shift occasionally over time. Bianchi *et al.* (2019) also document evidence of time-varying sparsity in factor models. We present the first approach (either Bayesian or frequentist) to do so.

The remainder of the paper is organized as follows. In Section 2 we detail our methodology. In Section 3 we present evidence of a simulation study. Section 4 introduces our empirical application, discussing evidence of breaks, the regime-specific selected factors, and their corresponding risk premia estimates. Section 5 discusses the pricing performance and investment implications of our selected factor collection and Section 6 concludes.

2. Methodology

We now outline the economic motivation for breaks in the risk factor model. Then, to build intuition, we explain how the methodology works for the no-break and single-break cases, before explaining our methodology for the most general case in which the subset of risk factors can shift across an unknown number of breaks that occur at unknown times. Finally, we detail the prior specification.

2.1. Economic Sources of Breaks in the Factor Model

Formally, suppose that for a time series sample from $t = 1, \dots, T$, we have data $\boldsymbol{f} = \{\boldsymbol{f}_t\}$, $t \leq T$ on a set of K (potential) risk factors. Suppose that the stochastic discount factor (SDF) at time t is given by

$$M_t = 1 - \boldsymbol{b}'(\boldsymbol{f}_t - \boldsymbol{\lambda})$$

where $\boldsymbol{b}$ is the vector of market prices of factor risks and $\boldsymbol{\lambda}$ is the vector of factor risk-premia. In an environment where the underlying firm-level production function is subject to breaks, due to technological innovations, it is more appropriate to assume that firm-level profitability would depend on a time-varying set of firm-level lagged characteristics. In this situation, the SDF would be more appropriately characterized by a time-varying SDF

$$M_t = 1 - \boldsymbol{b}_t'(\boldsymbol{f}_t - \boldsymbol{\lambda}_t)$$

where the market prices and factor risk-premia are time-varying. If we imagine that some of the lagged characteristics that determine firm-level profitability cease to be significant for periods of time due to changes in persistent shocks (innovations) to production, this would imply that some of the elements in the market price vector $\boldsymbol{b}_t$ would be zero and the corresponding elements of $\boldsymbol{f}_t$ would drop out of the SDF, that is, cease to be risk factors.

To describe this situation, let $\boldsymbol{x}_t \subseteq \boldsymbol{f}_t$ denote a subset of $\boldsymbol{f}_t$ with non-zero market prices of factor risks. Suppose that the market prices $\boldsymbol{b}_t$ change at unknown break dates

$$1 < t_1^* < t_2^* < \cdots < t_m^* < T \tag{1}$$

where m (the number of breaks) is also an unknown parameter. In particular, a different set of risk factors enters the SDF in each regime and thus there are $(m+1)$ risk factor sets

$$\boldsymbol{x}_t^* = \begin{cases} \boldsymbol{x}_t^1 & t \leq t_1^* \\ \boldsymbol{x}_t^2 & t_1^* < t \leq t_2^* \\ \vdots & \vdots \\ \boldsymbol{x}_t^m & t_{m-1}^* < t \leq t_m^* \\ \boldsymbol{x}_t^{m+1} & t_m^* < t \leq T. \end{cases} \tag{2}$$

The objectives of the analysis are to find

- the number of breaks $m \in \{0, 1, 2, \ldots, M\}$
- the timing of the breaks, $t_1^*, \ldots, t_m^*$
- the risk factors in each regime $\boldsymbol{x}_t^1, \ldots, \boldsymbol{x}_t^{m+1}$.

We now outline the framework developed by Chib and Zeng (2020) to find risk factors in the absence of breaks. We then generalize their framework to find risk factors with a single break in the market price vector (to help build intuition) and then consider the extension to multiple breaks (which we subsequently take to the data).

2.2. No breaks

Chib and Zeng (2020) develop a Bayesian model scanning approach to determine which subset of potential risk factors enters the SDF. To do this, they exploit the fact that asset pricing theory places restrictions on the joint distribution of factors that enter the SDF and those that do not. One key restriction is that the non-risk factors should be priced by the risk factors. One can therefore construct all possible decompositions of the joint distribution of factors in terms of a marginal distribution of the risk factors and a conditional distribution of the non-risk factors (imposing the pricing restriction on the latter) and determine by Bayesian marginal likelihoods, calculated by integrating out the parameters from the sampling density with respect to the prior of the parameters, which such decomposition is the best. The risk factors in that best decomposition are then taken to be the risk factors best supported by the data.

Therefore, to isolate the best set of risk factors, consider all possible splits of $\boldsymbol{f}_t$ into $\boldsymbol{x}_t$, the risk factors, and $\boldsymbol{y}_t$, the non-risk factors. These splits produce models that we indicate by $\mathbb{M}_j$, for $j = 1, \ldots, J = 2^K - 1$. At time t, the data-generating process under $\mathbb{M}_j$ is given by

$$\boldsymbol{x}_{j,t} = \boldsymbol{\lambda}_j + \boldsymbol{u}_{j,t}, \tag{3}$$

$$\boldsymbol{y}_{j,t} = \boldsymbol{\Gamma}_j \boldsymbol{x}_{j,t} + \boldsymbol{\varepsilon}_{j,t} \ , \qquad t = 1, \ldots, T \ , \tag{4}$$

where the errors are distributed as multivariate Gaussian:

$$\boldsymbol{u}_{j,t} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_j) \ , \ \boldsymbol{\varepsilon}_{j,t} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_j) \ . \tag{5}$$

Let the unknown parameters in this model be denoted by

$$\boldsymbol{\theta}_j = (\boldsymbol{\lambda}_j, \boldsymbol{\Omega}_j, \boldsymbol{\Gamma}_j, \boldsymbol{\Sigma}_j) \ , \ j \leq J \ . \tag{6}$$

Note that each of these models has a distinct set of risk factors and a distinct set of parameters.

Apart from $\boldsymbol{\lambda}_j$, the prior of the parameters $\boldsymbol{\Omega}_j, \boldsymbol{\Gamma}_j, \boldsymbol{\Sigma}_j$ are derived by change-of-variable from a *single* inverse Wishart prior placed on the matrix $\boldsymbol{\Omega}_j$ in the model where all factors are risk-factors. The hyperparameters of this single inverse Wishart distribution, and those of the model-specific $\boldsymbol{\lambda}_j$, are calculated from a training sample (which we take to be the first 15% of the sample data). The training sample data are subsequently discarded, and not used for estimation or model comparison purposes.

Let $\pi(\boldsymbol{\theta}_j)$ denote the prior on $\boldsymbol{\theta}_j$. Then, the marginal likelihood of $\boldsymbol{f}$ given $\mathbb{M}_j$ is

$$\text{marglik}(\boldsymbol{f}|\mathbb{M}_j) = \int \prod_{t=1}^{T} \mathcal{N}(\boldsymbol{x}_{j,t}|\boldsymbol{\lambda}_j, \boldsymbol{\Omega}_j)\mathcal{N}(\boldsymbol{y}_{j,t}|\boldsymbol{\Gamma}_j\boldsymbol{x}_{j,t}, \boldsymbol{\Sigma}_j)\, d\pi(\boldsymbol{\theta}_j), \quad j \leq J, \tag{7}$$

where $\mathcal{N}(\boldsymbol{z}|\boldsymbol{m}, \boldsymbol{V})$ is the multivariate normal density with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{V}$, evaluated at $\boldsymbol{z}$. These integrals are in closed form as shown in Chib *et al.* (2020). However, their approach assumes that the set of risk factors is time-invariant.

2.3. Single break

Assume for now the case of a single break. To find the unknown break date t_1^*, consider a break point t_1. For this break point, let the sample data on each side of the break be denoted by

$$\boldsymbol{f}_{s,1} = \{\boldsymbol{f}_t : t_{s-1} < t \leq t_s\}, \qquad s = 1, 2$$

where the second subscript in $\boldsymbol{f}$ indicates that this is the split in a setting with one break, and $s \in \{1, 2\}$ indicates the regime or segment s. A set of risk factors $\boldsymbol{x}_t^1$ enters the SDF in the first regime (from time periods $t = 1, \ldots, t_1$) and another set $\boldsymbol{x}_t^2$ enters in the second regime (from time periods $t = t_1 + 1, \ldots, T$). We assume that the risk factor set is stable within each regime.

To infer the optimal break date, and the identities of the risk factors $\boldsymbol{x}_t^1$ and $\boldsymbol{x}_t^2$, we focus on the quantity

$$\text{marglik}(\boldsymbol{f}_{1,1}, \boldsymbol{f}_{2,1}|t_1) \tag{8}$$

which is the marginal likelihood of the data segmented by the break date. We calculate this quantity on a large grid of possible break dates and let t_1^* be the date with the largest value of this marginal likelihood.

The problem in calculating the preceding quantity is that we do not know the identity of risk factors before and after the split. To deal with this two-way model uncertainty, we consider all possible divisions of $\boldsymbol{f}_t$ into $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$, on either side of t_1. Denote the models on the left of the split by $\mathbb{M}_{j_1,1}$ and on the right by $\mathbb{M}_{j_2,2}$, for $(j_1, j_2) = 1, ..., J = 2^K - 1$. When $j_1 = j_2$ the splits are identical but the parameters of the model are different. Just as we did in Equation (4), the jth model in regime s, $s = 1, 2$, for $t \in T_{s,1}$, where $T_{1,1} = \{t : t \leq t_1\}$ and $T_{2,1} = \{t : t > t_1\}$, takes the form

$$\begin{aligned}
\boldsymbol{x}_{j,s,t} &= \boldsymbol{\lambda}_{j,s} + \boldsymbol{u}_{j,s,t} \\
\boldsymbol{y}_{j,s,t} &= \boldsymbol{\Gamma}_{j,s}\boldsymbol{x}_{j,s,t} + \boldsymbol{\varepsilon}_{j,s,t} \\
\boldsymbol{u}_{j,s,t} &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}_{j,s}),\ \boldsymbol{\varepsilon}_{j,s,t} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{j,s}).
\end{aligned}$$

We denote the unknown parameters in these models by $\boldsymbol{\theta}_{j,s} = (\boldsymbol{\lambda}_{j,s}, \boldsymbol{\Omega}_{j,s}, \boldsymbol{\Gamma}_{j,s}, \boldsymbol{\Sigma}_{j,s})$. Note that each of these models has a different set of risk factors and a different set of parameters, and because we have a break, these parameters differ between regimes.

Letting $\pi(\boldsymbol{\theta}_{j,s})$ denote the prior on $\boldsymbol{\theta}_{j,s}$, the marginal likelihood of $\boldsymbol{f}_{s,1}$ given $\mathbb{M}_{j,s}$ and t_1, for $j \leq J$ and $s = 1, 2$, is

$$\text{marglik}(\boldsymbol{f}_{s,1}|\mathbb{M}_{j,s}, t_1) = \int \prod_{t \in T_{s,1}} \mathcal{N}(\boldsymbol{x}_{j,s,t}|\boldsymbol{\lambda}_{j,s}, \boldsymbol{\Omega}_{j,s}) \mathcal{N}(\boldsymbol{y}_{j,s,t}|\boldsymbol{\Gamma}_{j,s}\boldsymbol{x}_{j,s,t}, \boldsymbol{\Sigma}_{j,s})\, d\pi(\boldsymbol{\theta}_{j,s})$$

We can calculate these segment-wise marginal likelihoods by the method of Chib (1995).

Now by extending the argument and marginalization the marginal likelihood of the entire data given t_1 can be written as

$$\text{marglik}(\boldsymbol{f}_{1,1}, \boldsymbol{f}_{2,1}|t_1) = \sum_{j_1=1}^{J} \sum_{j_2=1}^{J} \text{marglik}(\boldsymbol{f}_{1,1}, \boldsymbol{f}_{2,1}|\mathbb{M}_{j_1,1}, \mathbb{M}_{j_2,2}, t_1) \Pr(\mathbb{M}_{j_1,1}) \Pr(\mathbb{M}_{j_2,2}) \quad (10)$$

$$= \frac{1}{J^2} \sum_{j_1=1}^{J} \sum_{j_2=1}^{J} \text{marglik}(\boldsymbol{f}_{1,1}|\mathbb{M}_{j_1,1}, t_1) \text{marglik}(\boldsymbol{f}_{2,1}|\mathbb{M}_{j_2,2}, t_1) \quad (11)$$

where in the second line we have assumed equal prior probabilities of models and the fact that the joint distribution factors into independent components given the models. In effect, what we do is pair each of the J possible models in the first regime with each possible model in the second and then marginalize over all possible such pairings.

We repeat the above calculation for every possible break date. The break date and two collections of regime-specific risk factors best supported by the data are those with the highest marginal likelihood.

2.4. Multiple breaks

With multiple breaks, we perform the same marginal likelihood calculation as in the single break approach, but this time, given m breaks, we calculate the marginal likelihood of the data segmented by the m breaks:

$$\text{marglik}(\boldsymbol{f}_{1,m}, ..., \boldsymbol{f}_{m+1,m}|t_1, \ldots, t_m). \quad (12)$$

We calculate this quantity for every possible combination of the m breaks and hence every possible combination of the J models in each of the $m + 1$ regimes.

Let the time points in the $(m + 1)$ regimes of $[1, T]$ induced by these m break dates be denoted by the sets

$$T_{s,m} = \{t : t_{s-1} < t \leq t_s\} \ , \qquad s = 1, ..., m+1. \tag{13}$$

Let the data on the factors in $T_{s,m}$ be given by

$$\boldsymbol{f}_{s,m} = \{\boldsymbol{f}_t : t_{s-1} < t \leq t_s\}, \qquad s = 1, ..., m+1. \tag{14}$$

Once again, we consider all possible splits of $\boldsymbol{f}_t$ into $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$, in each of the $m+1$ regimes. For regimes $s = 1, \ldots, m+1$, these splits produce models that we indicate by $\mathbb{M}_{j,s}$ for $j = 1, ..., J = 2^K - 1$. At time t, in regime s, the data generating process under $\mathbb{M}_{j,s}$ is given by

$$\begin{aligned}
\boldsymbol{x}_{j,s,t} &= \boldsymbol{\lambda}_{j,s} + \boldsymbol{u}_{j,s,t} \\
\boldsymbol{y}_{j,s,t} &= \boldsymbol{\Gamma}_{j,s}\boldsymbol{x}_{j,s,t} + \boldsymbol{\varepsilon}_{j,s,t} \\
\boldsymbol{u}_{j,s,t} &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}_{j,s}) \\
\boldsymbol{\varepsilon}_{j,s,t} &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{j,s}), \qquad t \in T_{s,m}.
\end{aligned}$$

Denoting the unknown parameters in these models by $\boldsymbol{\theta}_{j,s} = (\boldsymbol{\lambda}_{j,s}, \boldsymbol{\Omega}_{j,s}, \boldsymbol{\Gamma}_{j,s}, \boldsymbol{\Sigma}_{j,s})$, the marginal likelihood of $\boldsymbol{f}_{s,m}$ given $\mathbb{M}_{j,s}$ and the breaks $(t_1, \ldots, t_m)$ is given by

$$\text{marglik}(\boldsymbol{f}_{s,m}|\mathbb{M}_{j,s}, t_1, \ldots, t_m) = \int \prod_{t \in T_{s,m}} \mathcal{N}(\boldsymbol{x}_{j,s,t}|\boldsymbol{\lambda}_{j,s}, \boldsymbol{\Omega}_{j,s})\mathcal{N}(\boldsymbol{y}_{j,s,t}|\boldsymbol{\Gamma}_{j,s}\boldsymbol{x}_{j,s,t}, \boldsymbol{\Sigma}_{j,s})d\pi(\boldsymbol{\theta}_{j,s}) \tag{16}$$

The next step is to calculate the marginal likelihood of a given vector of change points given pairings of models from each of the $m+1$ regimes. There are $J^{(m+1)}$ such pairings in all regimes. For any one of these pairings, the marginal likelihood given the models in the pairings can be written as

$$\begin{aligned}
&\text{marglik}(\boldsymbol{f}_{1,m}, ..., \boldsymbol{f}_{m+1,m}|\mathbb{M}_{j_1,1}, \mathbb{M}_{j_2,2}, ..., \mathbb{M}_{j_{m+1},m+1}, t_1, \ldots, t_m) \\
&= \prod_{s=1}^{m+1} \text{marglik}(\boldsymbol{f}_{s,m}|\mathbb{M}_{j_s,s}, t_1, \ldots, t_m),
\end{aligned}$$

where we have used the fact that the joint factors into independent components given the models. We can get the desired marginal likelihood by summing the right hand side over all possible pairings of models. Specifically, we pair each of the J possible models in the first regime with each of the J possible models in each of the remaining m regimes and then marginalize over all possible such pairings. Thus,

$$\text{marglik}(\boldsymbol{f}_{1,m},...,\boldsymbol{f}_{m+1,m}|t_1,\ldots,t_m) = \frac{1}{J^{m+1}}\sum_{j_1=1}^{J}\cdots\sum_{j_{m+1}=1}^{J}\prod_{s=1}^{m+1}\text{marglik}(\boldsymbol{f}_{s,m}|\mathbb{M}_{j_s,s},t_1,\ldots,t_m). \tag{18}$$

This is the marginal likelihood for the break dates $t_1, \ldots, t_m$. The calculation is repeated for all possible locations of the m breaks. For this assumed number of m breaks, the optimal break dates $t_1^*, ..., t_m^*$ and the $m+1$ collection of regime-specific risk factors are those that have the highest marginal likelihood.

Finally, we repeat this calculation for different numbers of breaks $m \in \{0, 1, 2, ..., M\}$. The optimal number of breaks m, their corresponding break dates $(t_1^*, \ldots, t_m^*)$, and the set of risk factors selected in each of the $m+1$ regimes are those which have the highest marginal likelihood across all the number of breaks up to the maximum number considered. In the analysis we fix the maximum number of possible breaks, M, to be equal to three at the outset. This number depends on the sample size T. Due to the likely paucity of data within regimes, if M is too large relative to T, it is not realistic (or necessary) to have too many breaks. Moreover, since our Bayesian approach is based on the marginal likelihood which penalizes overparameterization, it will inherently guard against overfitting the number of breaks.[3]

2.5. Prior distributions

Our prior construction follows Chib and Zeng (2020). For completeness, we provide an abbreviated version here, but we refer the reader to their paper for full details. Formally, models in segment s are defined by the parameters

$$\boldsymbol{\theta}_{j,s} = (\boldsymbol{\lambda}_{j,s}, \boldsymbol{\psi}_{j,s})$$

where

$$\boldsymbol{\psi}_{j,s} = \begin{cases} \boldsymbol{\Omega}_{1,s} & j = 1 \\ (\boldsymbol{\Omega}_{j.s}, \boldsymbol{\Gamma}_{j,s}, \boldsymbol{\Sigma}_{j,s}) & j \geq 2. \end{cases}$$

The prior construction relies on the facts that the number of free parameters in

$$\boldsymbol{\psi}_{1,s} = \boldsymbol{\Omega}_{1,s},$$

[3]The identity in Chib (1995) notes that the log of the posterior ordinate is the penalty. Up to terms bounded in probability, this ordinate in large samples will equal log(T) multiplied by the number of parameters (Chib *et al.*, 2018). Thus, more complex models will have a bigger penalty.

is exactly equal to the number of free parameters in

$$\boldsymbol{\psi}_{j,s} \quad j \geq 2$$

and that there is a one-to-one mapping between them. Then, based on a training sample, which is common to all segments, a prior distribution is formulated for $\boldsymbol{\Omega}_{1,s}$. From this single distribution, by the change of variable, the distributions of $\boldsymbol{\psi}_{j,s}$, $j \geq 2$ are derived. In this way, the priors across models are equalized. This approach ensures consistency across model-specific priors, thereby enabling reliable model comparisons.

Finally, the prior of $\boldsymbol{\lambda}_{j,s}$ is also found from the same training sample common to all segments. This prior is conditioned on the error covariance matrix and is specified in such a way that it places little weight on implausibly large Sharpe ratios, thus avoiding approximate arbitrage opportunities (Pástor and Stambaugh, 1999).

2.6. Computational considerations

It is clear that the approach we have developed can be computationally intensive. With $K = 6$, there are $J = 63$ such splits, and with $m = 3$ change points, there are over 15 million model pairings that must be averaged for a single configuration of change points. This process must then be repeated for every possible configuration of change points. Though this burden is manageable for small to medium values of K, the burden increases rapidly with K. Then, practical trade-offs are required, such as limiting the analysis to one change point. Alternatively, one can adopt the framework of Chib *et al.* (2024) and work with the FF6 factors plus, say, the top five or seven principal components (PCs) of factors unexplained by the FF6 factors. In this way, one limits the initial pool of factors to a manageable number while accessing a wider set of factors through the PCs.

We now present the `R` function that performs the averaging over the different pairings of models. This function has been optimized for speed and is capable of handling billions of pairings efficiently. The input to the function is a list object of length $m + 1$, where each element in this list contains the marginal likelihoods for the J models in segment s.

```
meanprodlogm = function(logmargls = logmargls) {
    sumprod12 = function(x1,x2) {
    sapply(x1,FUN = function(x){return(sum(x*x2))})
    M1 = length(logmargls);
    M = M1 - 1;
    logmargm = do.call("cbind",logmargls);
    J = dim(logmargm)[1];
    xm = apply(logmargm,2,FUN = function(logmarg){
        thmax = max(logmarg);
```

```
        return(exp(logmarg-thmax))
    })
    thmax = apply(logmargm,2,FUN = function(logmarg){
        return(max(logmarg))
    })
    cind = M1;
    x3 = xm[,cind];
    x2 = xm[,(cind-1)];
    sp = sumprod12(x2,x3);
    j = M;
    while (j > 1) {
        x1 = xm[,(j-1)];
        sp123 = sumprod12(x1,sp);
        sp = sp123;
        j = j - 1;
    }
    smarg = sum(sp);
    logmarg = log(smarg) - M1*log(J) + sum(thmax);
    return(logmarg);
}
```

Listing 1: R function to calculate the log marginal likelihood for a given set of change points.

Another key ingredient into our approach is the set of potential change points:

$$\mathcal{C}_m = \Big\{(t_1, t_2, \ldots, t_m) \mid 1 < t_1 < t_2 < \cdots < t_m \leq T\Big\}.$$

In generating this set, it is advisable to impose the restriction that $t_i - t_{i-1} \geq n_0$, where n_0 represents some minimum segment length. The idea behind this constraint is to ensure reliable model estimation within each segment and also to reflect the belief that structural breaks occur relatively infrequently. We have written an efficient R function to generate this set, for $m \leq 4$. In one of our applications, where $T = 626$, $n_0 = 120$ and $m = 3$, the cardinality of $\mathcal{C}$ is 518665. Although this is a quite big set, one can use parallelized computations to efficiently handle the load.

```
makechangepoints = function(T, startdate, m, n0) {
    cfgs = list()
    idx = 1
    fl = T - n0 * m
    cp = numeric(m)
    for (s1 in seq(n0, fl)) {
        cp[1] = s1
        for (s2 in seq(s1 + n0, T - n0 * (m - 1))) {
```

```r
            cp[2] = s2
            if (m > 2) {
                for (s3 in seq(s2 + n0, T - n0 * (m - 2))) {
                    cp[3] = s3
                    if (m > 3) {
                        for (s4 in seq(s3 + n0, T - n0)) {
                            cp[4] = s4
                            cfgs[[idx]] = cp
                            idx = idx + 1
                        }
                    } else {
                        cfgs[[idx]] = cp
                        idx = idx + 1
                    }
                }
            } else {
                cfgs[[idx]] = cp
                idx = idx + 1
            }
        }
    }
    scfgs = do.call(rbind, cfgs)
    sl = apply(scfgs, 1, function(cp) c(cp[1], diff(cp), T - tail(cp, 1)))
    dts = seq(startdate, by = 1/12, length.out = T)
    pdts = as.character(format(as.yearmon(dts), "%b %Y"))
    cpls = lapply(seq_len(nrow(scfgs)), function(i) pdts[scfgs[i, ]])
    names(cpls) = sapply(cpls, function(e) paste(e, collapse = " | "))
    list(cpls = cpls, nsls = split(t(sl), seq_len(ncol(sl))))
}
```

Listing 2: R function to create a list of possible change points.

3. Simulation study

We conduct two sampling experiments. For realism, each experiment is matched to the real data that are analyzed in the next section. Those data consist of 726 monthly observations on the FF6 factors. The first 100 of those observations are taken as the training sample for the prior. This sample is fixed in the sampling experiments and is not simulated. On the remaining 626 observations we run our change point test, first with one change point and then we two change points and use the best fitted models to simulate 100 data sets.

3.1. Design 1

This is with a single change point. On the real data, according to our test, the break occurs at May 1998. The best fitted models at the posterior mean of the parameters are as follows. On the left of the split (the first segment),

$$\begin{bmatrix} \text{Mkt} \\ \text{SMB} \\ \text{HML} \\ \text{RMW} \\ \text{CMA} \\ \text{MOM} \end{bmatrix} = \begin{bmatrix} 0.006 \\ 0.0015 \\ 0.005 \\ 0.0026 \\ 0.0034 \\ 0.0088 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_6 \end{bmatrix}$$

where $\boldsymbol{u} \sim N(0, \boldsymbol{L}_1\boldsymbol{L}_1')$ and

$$\boldsymbol{L}_1 = \begin{bmatrix} 0.0455 & 0 & 0 & 0 & 0 & 0 \\ 0.0069 & 0.0281 & 0 & 0 & 0 & 0 \\ -0.0109 & 0.0014 & 0.0244 & 0 & 0 & 0 \\ 0.0012 & -0.0041 & -0.0065 & 0.0131 & 0 & 0 \\ -0.0081 & 2e-04 & 0.0111 & -0.0029 & 0.0108 & 0 \\ 0.0011 & -0.0062 & -0.005 & 0.0015 & 0.005 & 0.0322 \end{bmatrix}$$

To the right of the split (the second segment), the best model at the posterior mean of the parameters is

$$\begin{bmatrix} \text{Mkt} \\ \text{SMB} \\ \text{HML} \\ \text{RMW} \\ \text{CMA} \end{bmatrix} = \begin{bmatrix} 0.006 \\ 0.0021 \\ 7e-04 \\ 0.0035 \\ 0.0022 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_5 \end{bmatrix}$$

where now

$$\boldsymbol{L}_2 = \begin{bmatrix} 0.0475 & 0 & 0 & 0 & 0 \\ 0.0086 & 0.031 & 0 & 0 & 0 \\ -0.0024 & 0.0024 & 0.0354 & 0 & 0 \\ -0.0107 & -0.0115 & 0.0111 & 0.0225 & 0 \\ -0.0068 & 0.0022 & 0.0144 & -0.0011 & 0.0171 \end{bmatrix}$$

and

$$\begin{bmatrix}\text{MOM}\end{bmatrix} = \mathbf{\Gamma}\begin{bmatrix}\text{Mkt}\\ \text{SMB}\\ \text{HML}\\ \text{RMW}\\ \text{CMA}\end{bmatrix} + \varepsilon$$

where

$$\mathbf{\Gamma} = \begin{bmatrix}-0.3296 & 0.1865 & -0.5922 & 0.1901 & 0.4024\end{bmatrix}$$

and $\varepsilon \sim N(0, 0.047^2)$. Thus, in this design, the risk factor set goes from six to five. The parameters of the models also change, but by relatively small amounts. We have written out the models to ease replication, but also to highlight the difficult problem at hand. We do not know where this break occurs and we do not know which of the 63 models is generating the data to the left of any possible split or to the right of the split.

We generate 100 data sets of length 626 from Design 1. For each data set, we let $\mathcal{C}_1$ consist of 387 months (starting from October 1981 and going to December 2013) and we record the change point detected by our marginal likelihood based test. Then across the 1000 data sets, we calculate the empirical distribution of these detected change points. This distribution is exhibited in Figure 1. The empirical distribution concentrates on the true value of the change point, demonstrating the effectiveness of the proposed test.

3.2. Design 2

In this design we consider two change points. The timing of the change points and the parameters of the model are matched to the data. In the first segment, which runs from August 1971 to July 1998, we have six risk factors generated according to

$$\begin{bmatrix}\text{Mkt}\\ \text{SMB}\\ \text{HML}\\ \text{RMW}\\ \text{CMA}\\ \text{MOM}\end{bmatrix} = \begin{bmatrix}0.006\\ 0.0013\\ 0.0049\\ 0.0026\\ 0.0033\\ 0.0091\end{bmatrix} + \begin{bmatrix}u_1\\ u_2\\ \vdots\\ u_6\end{bmatrix}$$

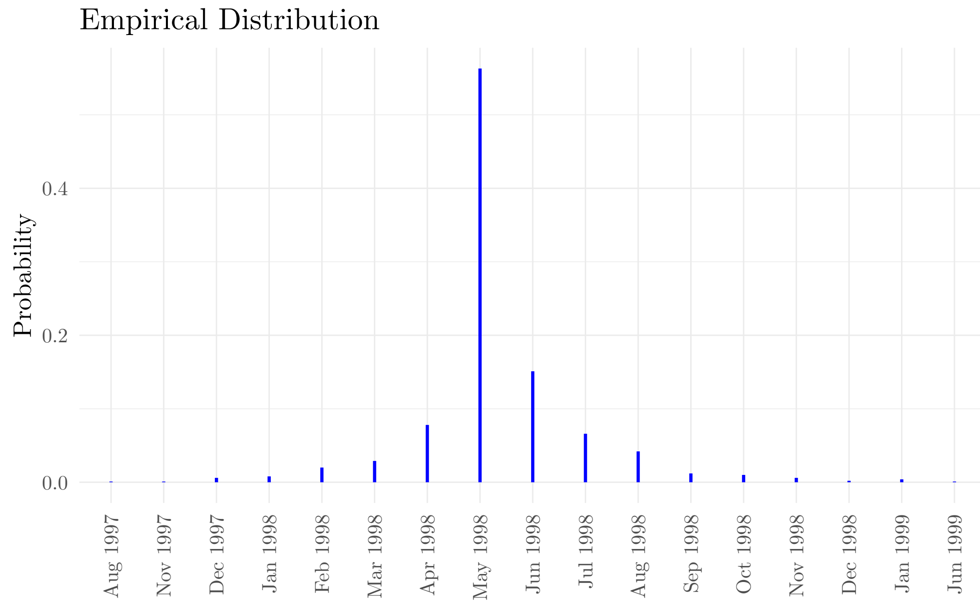

Figure 1: Design 1: Empirical distribution of detected change points, based on 100 simulated data sets and a set of potential change points $\mathcal{C}_1$ consisting of 387 months. The empirical distribution is plotted for the months with non-zero empirical probability. The true change point occurs at May 1998. The empirical distribution concentrates at this point.

and the lower triangular cholesky factor of $\boldsymbol{\Omega}_1$ is

$$\boldsymbol{L}_1 = \begin{bmatrix} 0.0454 & 0 & 0 & 0 & 0 & 0 \\ 0.0069 & 0.0282 & 0 & 0 & 0 & 0 \\ -0.0109 & 0.0016 & 0.0244 & 0 & 0 & 0 \\ 0.0011 & -0.0042 & -0.0065 & 0.0131 & 0 & 0 \\ -0.0081 & 3e-04 & 0.0111 & -0.0028 & 0.0109 & 0 \\ 0.0012 & -0.0066 & -0.0051 & 0.0013 & 0.0046 & 0.0323 \end{bmatrix}$$

Then, in the second segment, that runs from August 1998 to March 2009, we have two risk factors generated according to the model

$$\begin{bmatrix} \text{SMB} \\ \text{RMW} \end{bmatrix} = \begin{bmatrix} 0.0046 \\ 0.0042 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

where

$$\boldsymbol{L}_2 = \begin{bmatrix} 0.0377 & 0 \\ -0.0213 & 0.0338 \end{bmatrix}$$

and four non risk factors generated as

$$\begin{bmatrix} \text{Mkt} \\ \text{HML} \\ \text{CMA} \\ \text{MOM} \end{bmatrix} = \begin{bmatrix} -0.174 & -0.7817 \\ 0.1641 & 0.6673 \\ 0.1649 & 0.3294 \\ 0.5389 & 0.3838 \end{bmatrix} \begin{bmatrix} \text{SMB} \\ \text{RMW} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

and the lower triangular cholesky factor of $\boldsymbol{\Sigma}_2$ is

$$\boldsymbol{L}_{2,y.x} = \begin{bmatrix} 0.0417 & 0 & 0 & 0 \\ 0.0037 & 0.0299 & 0 & 0 \\ -0.0075 & 0.0138 & 0.0186 & 0 \\ -0.0165 & -0.0149 & 0.0025 & 0.0539 \end{bmatrix}$$

Finally, in the third segment, going from April 2009 to December 2023, we again have two risk factors. The data generating process for the risk factors is

$$\begin{bmatrix} \text{Mkt} \\ \text{RMW} \end{bmatrix} = \begin{bmatrix} 0.0112 \\ 0.0027 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\boldsymbol{L}_3 = \begin{bmatrix} 0.0454 & 0 \\ -0.0019 & 0.0201 \end{bmatrix}$$

and for the non risk factors it is

$$\begin{bmatrix} \text{SMB} \\ \text{HML} \\ \text{CMA} \\ \text{MOM} \end{bmatrix} = \begin{bmatrix} 0.2089 & -0.5027 \\ 0.0856 & 0.0502 \\ -0.0381 & 0.1028 \\ -0.3631 & -0.0134 \end{bmatrix} \begin{bmatrix} \text{Mkt} \\ \text{RMW} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

$$\boldsymbol{L}_{3,y.x} = \begin{bmatrix} 0.0236 & 0 & 0 & 0 \\ 0.0107 & 0.0318 & 0 & 0 \\ 0.0032 & 0.0138 & 0.0156 & 0 \\ -0.0099 & -0.0101 & 0.01 & 0.0395 \end{bmatrix}$$

We now generate 100 data sets of length 626 from Design 2. For each data set, the set of potential change points $\mathcal{C}_2$ consists of 35,778 pairs of months (starting from October 1981 and going to December 2013). The optimal change point is found by applying our test to each of these possible change points and selecting the one with the highest marginal likelihood. This test is repeated for each of the 100 simulated data sets. The empirical distribution of the detected change points for pairs of points with non-zero empirical probability is given in Figure 2. The empirical distribution concentrates on the pair July 1998 and February 2009. The first change point in this pair is correct and the second is off by a month. Given that the second change point is during the period of the 2008-2009 financial crisis, and Design 2 is matched to the data at that time, it is not surprising that the precise change point during that crisis is difficult to narrow down. Nonetheless, the ability of the test to get this close to the true value is remarkable.

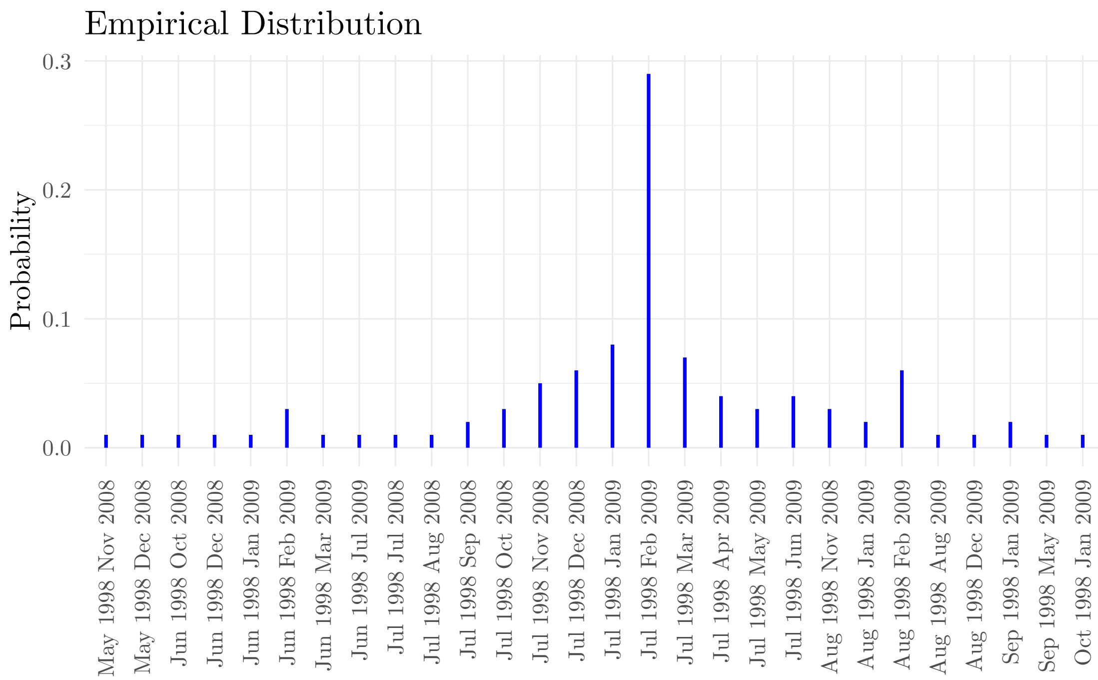

Figure 2: Design 2: Empirical distribution of detected change points, based on 100 simulated data sets and a set of potential change points $\mathcal{C}_2$ consisting of 35778 pairs of months. The empirical distribution is plotted for the subset of 35778 pairs of months with non-zero empirical probability. The true change point occurs at July 1998 and March 2009. The empirical distribution concentrates correctly on the first change point and is off by a month for the second change point.

4. Factor selection

This section first describes our data before detailing our results on evidence of breaks and regime-specific factor selection. Finally, we discuss implications for estimates of time-varying risk premia and market prices of risk.

4.1. Data

Our analysis centers on the six factors introduced by Fama and French (2018), with detailed descriptions provided in Table 1. The data, spanning from August 1963 to December 2023, is sourced from Ken French's website and consists of 726 monthly observations. Of these, the initial 100 months are designated as the training sample to construct the prior for each segment. The remaining 626 months are utilized for estimation and inference.

Table 1: **Definitions of factors used in the study**. This Table defines how the factors used in our study are constructed.

Factors	Definitions
MKT	the excess return of the market portfolio
SMB	the return spread between diversified portfolios of small size and big size stocks
HML	the return spread between diversified portfolios of high and low B/M stocks
RMW	the return spread between diversified portfolios of stocks with robust and weak profitability
CMA	the return spread between diversified portfolios of stocks of low (conservative) and high (aggressive) investment firms
MOM	the return spread between diversified portfolios of stocks with high and low returns over the previous 12 months

4.2. Rolling window

Bessembinder *et al.* (2021) also consider time variation in the factor zoo. They select a factor if the intercept of an OLS regression of the factor on the MKT factor has a t-statistic greater than 3.0. Then, they apply this using a fixed 60-month rolling window approach throughout their sample to select factors over time.

Although there may be occasional shifts in the set of risk factors that explain the cross section of expected returns due to, for example, changes in monetary policy, regulation, or technological innovations, having factors enter and exit the set of true factors every month is difficult to motivate from an economic standpoint. The rolling window causes factors to enter and exit the selected set in a noisy manner, which is difficult to motivate economically.

To see how the rolling window approach causes factors to enter and exit the set in a noisy fashion, each window of Figure 3 displays whether the factor is selected (indicator variable equals one) or omitted (indicator variable equals zero) at a given point in time, from the best model using the Chib and Zeng (2020) scan on data from rolling windows of length 60 months. There are 567 such rolling windows from Nov 1971 to Dec 2023. Forty eight of those 567 windows are labeled on the x-axis. If one uses these rolling window estimates to calculate the corresponding total number of risk factors selected for each interval, one sees that the number of factors selected in the model varies between one and six and this number

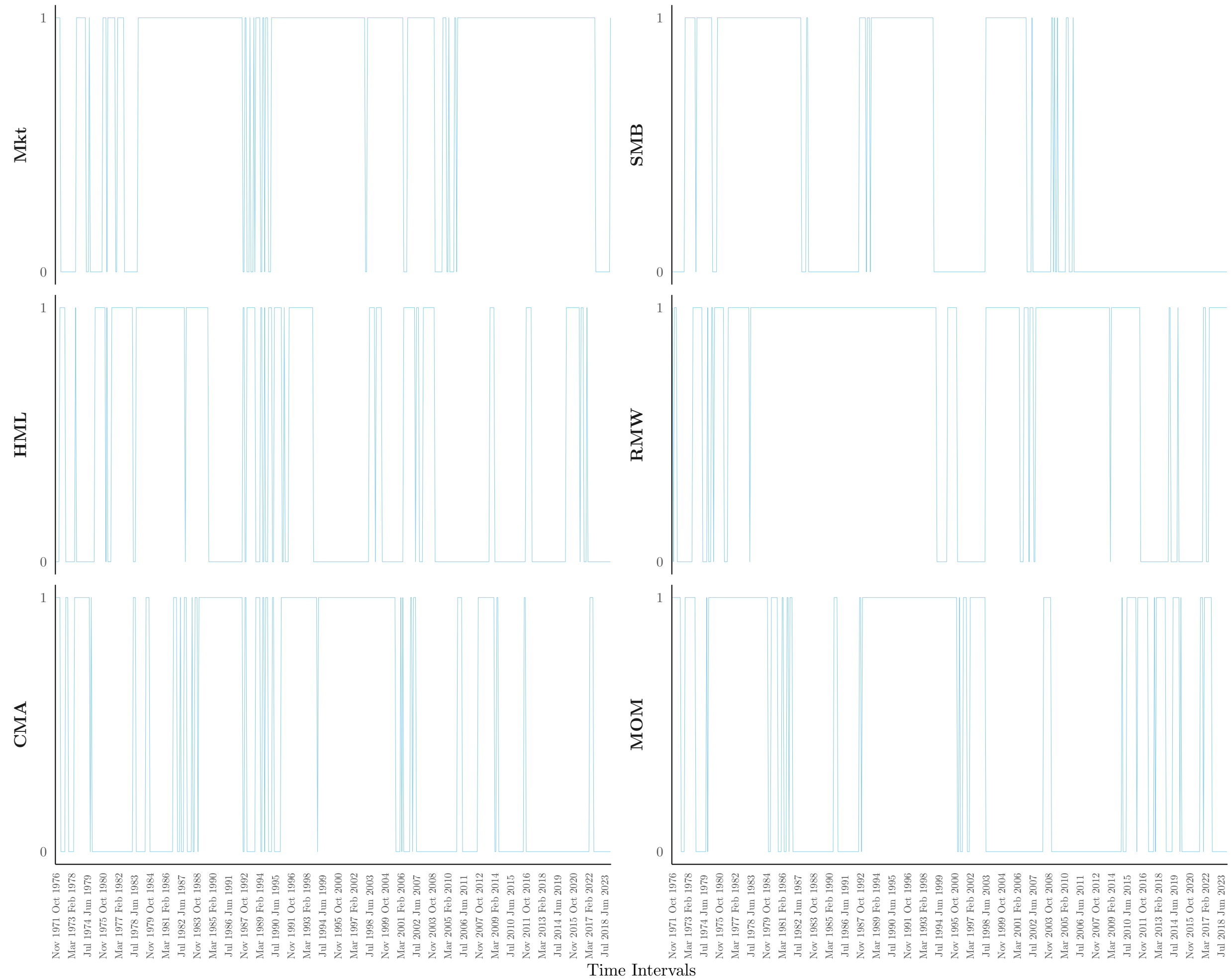

Figure 3: Each subfigure displays whether the factor is selected (indicator variable equals one) or omitted (indicator variable equals zero) at a given point in time, from the best model using the Chib and Zeng (2020) scan on data from rolling windows of length 60 months. There are 567 such rolling windows from August 1971 to December 2023. Forty eight of those 567 windows are labeled on the x-axis.

is changing very frequently. Such frequent changes in the risk factor set are hard to motivate on economic grounds.

This evidence can be used as motivation for our methodology, which restricts the number of shifts in the set of risk factors, assuming that the set is stable between structural breaks.

4.3. Evidence of breaks

We now determine the evidence in support for single, double and triple breaks. We construct the sets $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$ using Listing 2 with $T = 626$, startdate of Nov 1971 and $n_0 = 60$. The cardinality of these sets is 387, 35778 and 518665, respectively. We could also consider $m = 4$ break points and would likely detect one around the time of the Covid-Pandemic, but that would leave us with too few observations in the last segment to do valid estimation of the different models. For data paucity reasons, therefore, we only consider up to 3 break points.

The top panel of Table 2 displays the log marginal likelihoods for the optimal break dates when assuming different numbers of breaks from zero to three. We see that three breaks

Table 2: **Number of breaks:** Log marginal likelihoods under different numbers of breaks and their locations estimated using our methodology on the six factor model of Fama and French (2018) using data from Nov 1971 through Dec 2023. The row displayed in bold font corresponds to the optimal number and timing of breaks.

No. of breaks	Log marg lik	Break dates		
0	8022.578			
1	8223.860	May 1998		
2	8312.709	Jul 1998	Mar 2009	
3	**8349.452**	**Mar 1982**	**Jul 1998**	**Mar 2009**
3	8349.449	Nov 1982	Jul 1998	Mar 2009
3	8349.135	Feb 1982	Jul 1998	Mar 2009
3	8348.925	Oct 1982	Jul 1998	Mar 2009
3	8348.898	Apr 1982	Jul 1998	Mar 2009
3	8348.844	May 1982	Jul 1998	Mar 2009

have the highest logarithmic marginal likelihood and, therefore, are clearly preferred to fewer numbers of breaks. The optimal timing for these three breaks is March 1982, July 1998, and March 2009. Interestingly, the identity of the first change point stays almost intact when we go two change points and the identity of those change points stays almost the same when we bring in the third. This shows that the change points we identify are robust to the number of change points.

The break dates we find correspond closely to major events such as the end of the 1979-1982 “monetarist policy experiment” implemented by Paul Volcker following the oil price shocks of the 1970s, the rise of the Internet revolution, and digitization of financial markets that culminated in the dotcom bubble and subsequent bursting, and the Global Financial Crisis. The post-2009 data correspond to the current regime and thus these are the data that are relevant for finding risk factors that are currently pricing the cross section.

The bottom panel of Table 2 displays the log marginal likelihood for three break dates that are very close to the optimal three break dates. We see in each case that the log marginal likelihood is slightly lower, and therefore these break dates are dominated by the optimal ones. Figure 4 illustrates the posterior probabilities associated with the top 30 break date combinations, given the preferred three-break model. The results show a clear preference for the optimal break dates, which capture nearly 15 percent of the total posterior probability. The second most likely combination, which shifts the first break from March 1982 to November 1982, also secures nearly 15 percent of the posterior probability. The probabilities decline sharply thereafter, with the eighth most likely break date combination capturing just five percent of the total posterior weight.

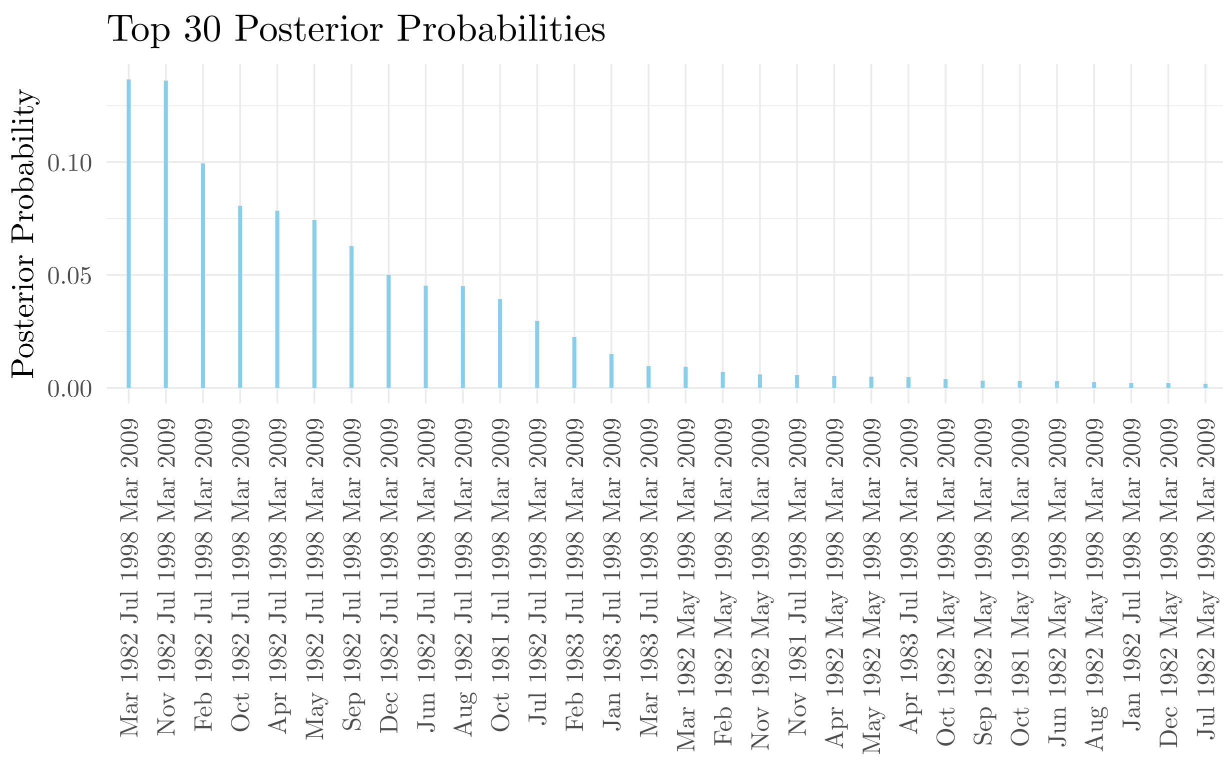

Figure 4: This figure displays the posterior probability assigned to each of the top 30 break date combinations for the preferred three break model. These are the top 30 of 540274 three break combinations considered.

4.3.1. Instant or gradual changes?

Occasional changes in the set of risk factors could be driven by publication effects (McLean and Pontiff, 2016) or important regulatory or technological changes. These types of events are likely to cause the set of risk factors to shift abruptly. Alternatively, risk and risk premia change more gradually over time, either as a function of business conditions or as a result of slow-moving changes in the economy. Given that the nature of the change is unobservable, it is important to incorporate both potential changes, abrupt and gradual, when estimating changes in the set of risk factors.

Although we focus on inference in this paper on the optimal break date, an important feature of our Bayesian framework is that the posterior distribution over change points reveals the uncertainty surrounding the break dates. The uncertainty surrounding the break date estimate can be incorporated into risk premia and price of risk estimates from the factor model.

4.4. Risk factor selection

The first four panels of Table 3 show log marginal likelihoods for the top five models ranked by log marginal likelihood in the four regimes separated by the optimal three break dates. A value equal to one (zero) indicates that a factor is selected (omitted). In the first regime (1971-1982), the optimal model is the one which includes all six potential factors except the Market factor. This collection of risk factors has a log marginal likelihood of 1695.916. Three of the top five models in this regime omits the market factor, providing reasonably strong evidence that this factor is not relevant in this regime. The only two factors that are selected in each of the five top performing models are the profitability factor and the momentum factor. The upper left window of Figure 5 displays, for this regime, the log marginal likelihoods for each of the 63 possible models ranked based on the log marginal likelihood from best (left) to worst (right). The best model, which selects all six factors except MKT, is shown in blue. Other models of interest are also colored blue. We see that dense models that include more factors (FF6) tend to perform better than sparse models (FF3).

The upper left window of Figure 6 displays, for this regime, the number of factors in each of the 63 possible models that are ranked based on log marginal likelihood from best (left) to worst (right). The best model along with some notable Fama-French models are colored blue. The remaining models are colored red. We see that the best performing models (to the left of the figure) tend to be dense models (toward the top), while the worst performing models (to the right) tend to be sparse models (toward the bottom). This finding suggests that dense factor models are preferred to sparse models during this period.

In the second regime (1982-1998), the FF6 model is the best performing model with a log marginal likelihood of 2863.626. Each of the five best-performing models includes at least five factors. The profitability factor is again selected in each of the top five models, as is the market factor. Once again, popular dense models tend to perform well (FF6 and FF5), while popular sparse models (FF3) tend to perform poorly. In fact, the pattern in which dense models tend to outperform sparse models is even more striking in this regime (top right window of Figure 6).

This patterns changes markedly in the final two regimes (post-1998). Here, the preferred model contains just two factors: size and profitability until 2009, after which it is profitability and market. The size factor is clearly omitted after 2009. The value and momentum factors are rarely selected in either of these two regimes.

These findings make a simple, yet novel point. Until 1998, dense models were preferred,

Table 3: **Regime-specific factor selection:** The first four panels of this table display log marginal likelihoods for the top five models ranked by log marginal likelihood in the four regimes separated by the optimal three break dates. A value equal to one (zero) indicates a factor is selected (omitted). The final panel displays the same information from the model that precludes breaks.

Mkt	SMB	HML	RMW	CMA	MOM	Log marg lik
Nov 1971 - Mar 1982						
0	1	1	1	1	1	1695.916
0	1	1	1	0	1	1695.901
1	1	1	1	1	1	1695.563
1	0	1	1	1	1	1695.547
0	1	0	1	1	1	1695.525
Apr 1982 - Jul 1998						
1	1	1	1	1	1	2863.626
1	1	1	1	0	1	2861.870
1	0	1	1	1	1	2861.860
1	1	1	1	1	0	2861.667
1	1	0	1	1	1	2861.442
Aug 1998 - Mar 2009						
0	1	0	1	0	0	1483.328
1	1	1	1	1	0	1483.132
1	1	0	1	0	0	1482.892
0	1	0	1	1	0	1482.567
0	1	0	1	0	1	1482.518
Apr 2009 - Dec 2023						
1	0	0	1	0	0	2316.203
1	0	0	1	0	1	2315.900
1	0	0	1	1	0	2315.282
1	0	0	0	0	0	2315.235
1	0	1	1	1	0	2315.141
Nov 1971 - Dec 2023 (No breaks)						
1	0	0	1	1	1	8022.578
1	1	0	1	1	1	8022.235
1	0	1	1	1	1	8020.841
1	1	1	1	1	1	8020.491
1	0	0	0	1	1	8016.994

but since then dense models do not outperform sparse models and, in fact, the best model in both of the final two regimes selects just two factors. Ignoring breaks conceals this changing dynamic and would lead one to spuriously believe that dense models are still preferable

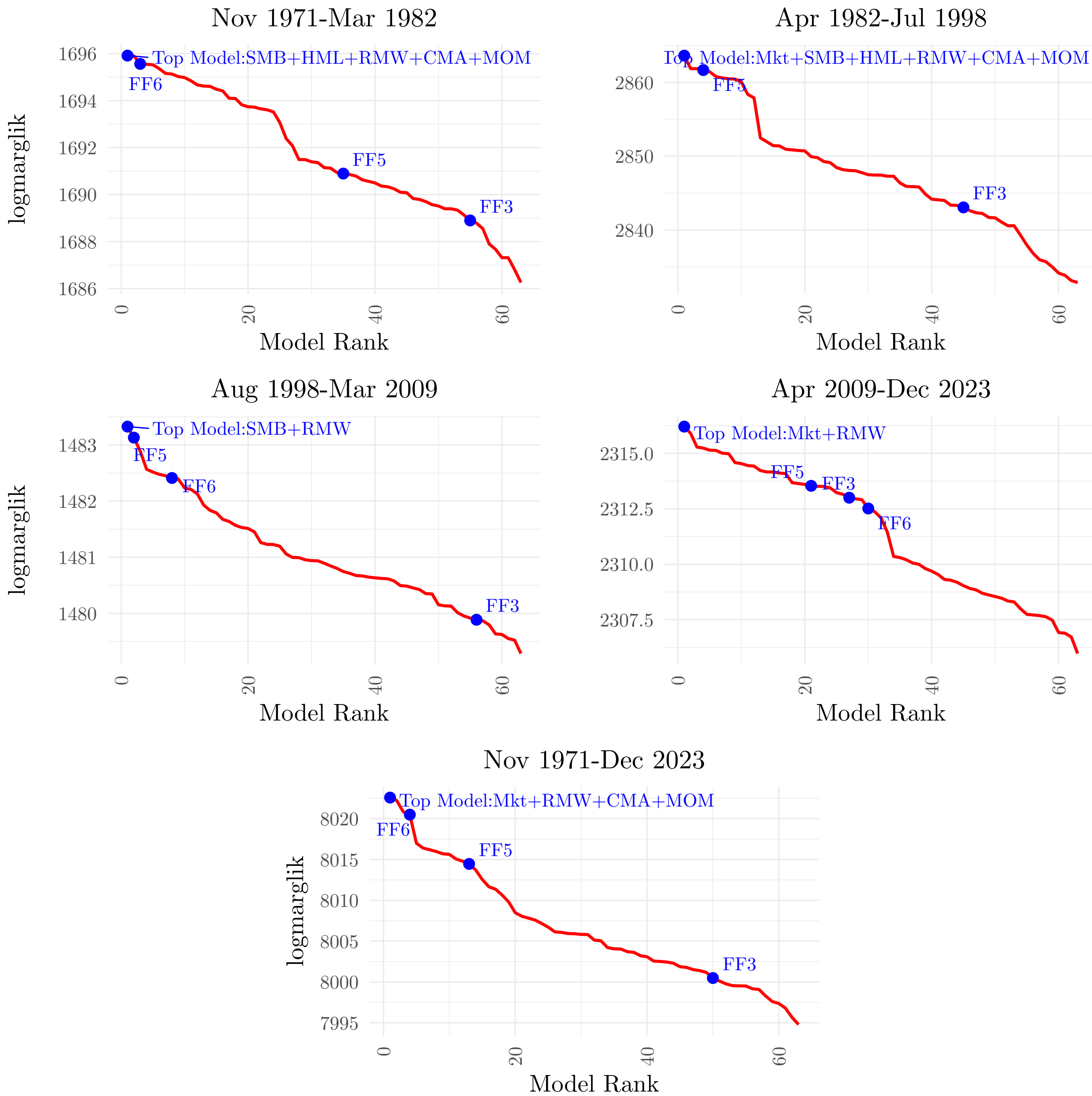

Figure 5: The top four windows of this Figure display, for each of the four regimes, the log marginal likelihoods for each of the 63 possible models which are ranked based on log marginal likelihood from best (left) to worst (right). The best model (top-left circle in each panel) and the Fama-French models are colored blue. The lower panel displays the same information for the model that precludes breaks.

today. This is because using pre-break data tends to detect factors that were once relevant but not at present. The implication of this overlooked finding is that there has been a clear shift toward parsimony, and researchers should avoid using pre-break data when selecting factors. Some recent studies that identify a large number of risk factors may in part reflect this phenomenon: several factors in those models may simply be fitting prebreak data that are no longer relevant. We recommend that researchers use our approach or use only the

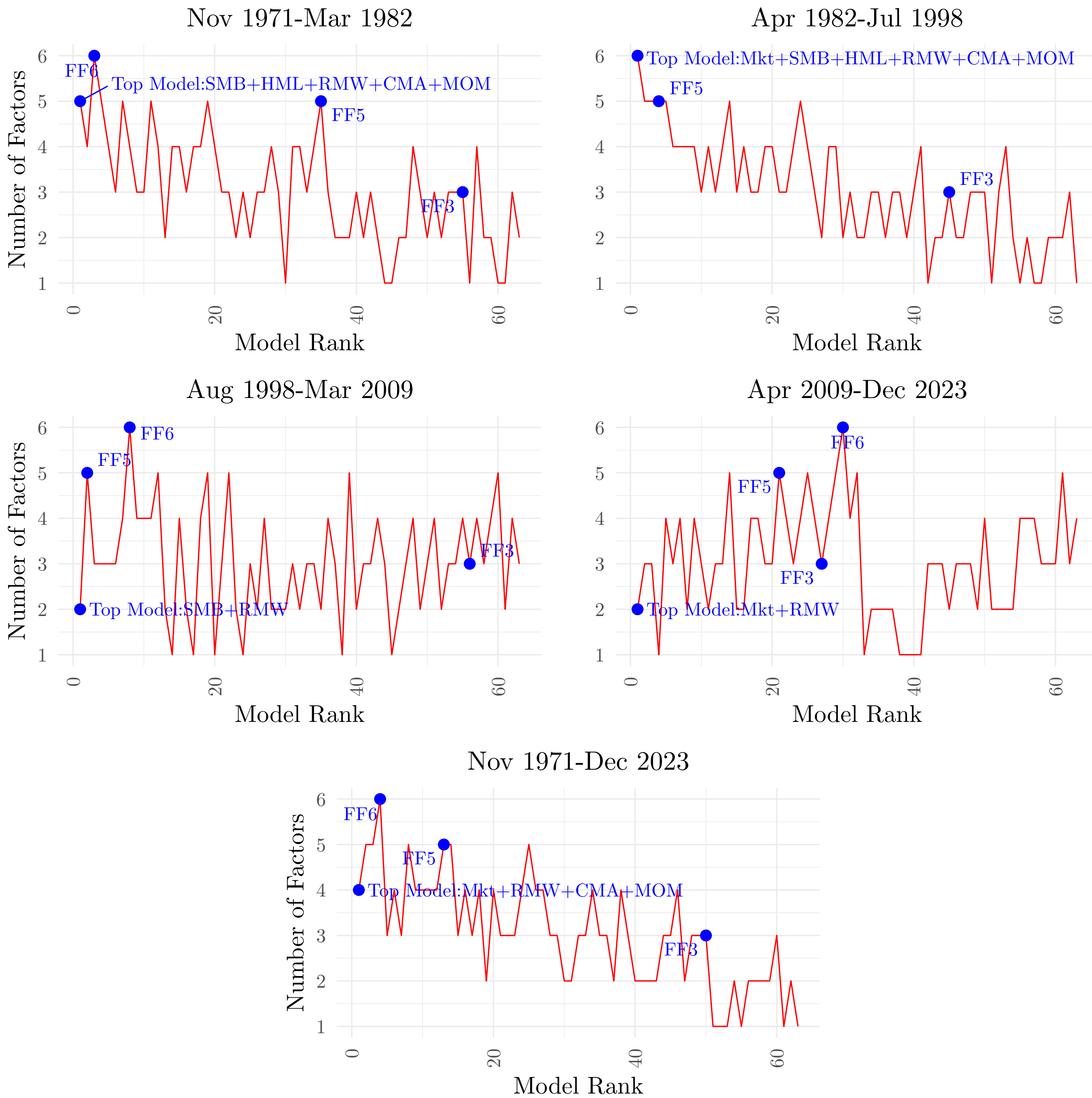

Figure 6: The top four windows of this Figure display, for each of the four regimes, the number of factors in each of the 63 possible models which are ranked based on log marginal likelihood from best (left) to worst (right). The best model and the Fama-French models are displayed in blue. The lower panel displays the same information for the model that precludes breaks.

post-1998 data in their future analyses.

4.5. Factor risk premia and market prices of risk

Our approach also generates estimates of factor risk premia and their market prices of risk, which are allowed to vary across regimes.

The four top panels of Table 4 present estimates of the risk premia of the factors selected

Table 4: **Risk premia estimates**. The five panels of this table display, for each of the four regimes and the full period, a range of risk premia estimates for the factors selected by the optimal model in that regime. Specifically, we report the posterior mean and standard deviation of the risk premia estimates. We also report the lower and upper estimates that correspond to the 95 percentiles of the posterior distribution.

	Mkt	SMB	HML	RMW	CMA	MOM
			Nov 1971 - Mar 1982			
postmean		0.0053	0.0061	-0.0003	0.0039	0.0103
postsd		0.0029	0.0025	0.0015	0.0016	0.0035
lower		-0.0003	0.0013	-0.0032	0.0007	0.0034
upper		0.0110	0.0112	0.0026	0.0070	0.0172
			Apr 1982 - Jul 1998			
postmean	0.0092	-0.0010	0.0039	0.0044	0.0028	0.0081
postsd	0.0029	0.0018	0.0017	0.0009	0.0012	0.0019
lower	0.0034	-0.0045	0.0006	0.0026	0.0005	0.0043
upper	0.0149	0.0024	0.0071	0.0061	0.0052	0.0119
			Aug 1998 - Mar 2009			
postmean		0.0046		0.0042		
postsd		0.0031		0.0033		
lower		-0.0016		-0.0022		
upper		0.0107		0.0108		
			Apr 2009 - Dec 2023			
postmean	0.0112			0.0027		
postsd	0.0032			0.0014		
lower	0.0048			-0.0001		
upper	0.0175			0.0056		
			Nov 1971 - Dec 2023			
postmean	0.0060			0.0031	0.0029	0.0059
postsd	0.0018			0.0009	0.0008	0.0017
lower	0.0025			0.0013	0.0013	0.0026
upper	0.0096			0.0049	0.0044	0.0093

by the optimal model in each of the four regimes. Specifically, we report the posterior mean and standard deviation of the risk premia estimates, along with the lower and upper bounds corresponding to the 95th percentiles of the posterior distribution. The bottom panel displays the same information for the model that precludes breaks. Figure 7 illustrates the corresponding estimated posterior densities of the risk premia.

We see that the equity premium is estimated to be about seven percent annualized when precluding breaks. Accounting for breaks, however, induces some time variation around this value, with the equity premium generally rising throughout the sample and reaching its

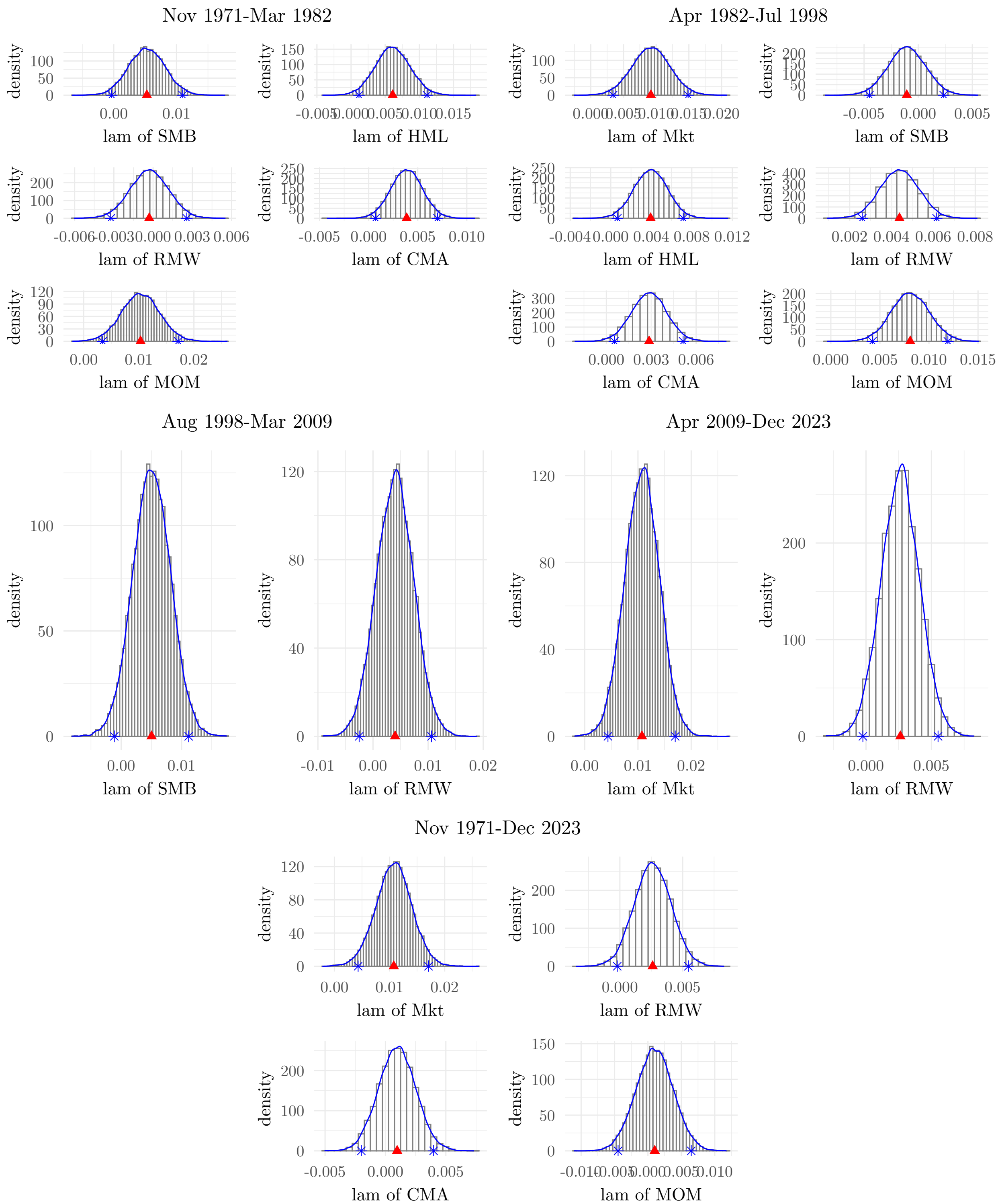

Figure 7: The top four windows of this Figure display, for each of the four regimes, the estimated posterior density risk premia plots for each of the selected risk factors in the optimal model. The bottom panel displays the same information for the model that precludes breaks.

highest value in the most recent regime (post-1998).

The only factor consistently selected throughout the sample and in the current regime is the profitability factor. The estimated premium on this factor has shown significant time variation, rising from a low value to approximately 5% between 1982 and 2009, before moderating somewhat. In contrast, the no-break model smooths through these regime changes, estimating a steady premium of around 3.5% throughout the sample period.

We also present clear evidence of a decrease in the value premium over time, aligning with findings documented in previous studies (Fama and French, 2021). Specifically, we observe that the value premium drops from about seven percent prior to 1982 to about five percent from 1982 through 1998. Since 1998, the value factor has been omitted altogether.

The size factor has declined over time and is omitted altogether in the final regime. In the most recent regime, the market and profitability factors are the only two selected, both generating significant risk premia. In summary, we find evidence of time variation in all six-factor risk premia.

Regarding the market prices of the factor risks, which are the weights of the risk factors in the SDF, Table 5 shows clear evidence of time variation in the market price of risk for the six factors that is obscured when breaks are excluded.

5. Pricing performance and investment implications

5.1. Pricing of the excluded factors

Having identified optimal regime-specific collections of risk factors, we next investigate whether the selected factors in each regime can price the omitted ones. To evaluate pricing ability, we fit a series of regression models with each excluded factor as the dependent variable and the selected risk factors as the independent variables. This analysis is conducted separately for each regime, using only the data and factors selected from that regime. For each excluded factor, two Bayesian regressions are estimated: one with an intercept and one without. The marginal likelihood of each model is then calculated using the method of **?**.

If the log marginal likelihood of the model without an intercept exceeds that of the model with an intercept by more than 0.69, then we can conclude that the omitted factor is priced by the selected factors in that regime. Formally, exceeding this threshold indicates that the posterior odds of the model without an intercept relative to the model with an intercept are at least 2:1.[4] The results of applying this test to each of the excluded factors are shown in Table 6.

[4]Frequentist tests are based on sampling distributions of estimators, which require the involvement of unseen samples beyond the one that is observed. The Bayesian test is only conditioned on the observed data. In addition, the Bayesian test is based on the estimation of both models, not just one model as in a frequentist

Table 5: **Price of risk estimates**. The five panels of this table display, for each of the four regimes and the full period, a range of market price of risk estimates for the factors selected by the optimal model in that regime. Specifically, we report the posterior mean and standard deviation of the risk premia estimates. We also report the lower and upper estimates that correspond to the 95 percentiles of the posterior distribution.

	Mkt	SMB	HML	RMW	CMA	MOM
Nov 1971 - Mar 1982						
postmean		6.5750	8.9706	17.9286	12.1548	8.1802
postsd		3.0189	5.4385	8.1384	8.6793	2.4963
lower		0.7358	-1.6744	2.4894	-4.4251	3.4067
upper		12.6128	19.8437	34.4599	29.4631	13.1993
Apr 1982 - Jul 1998						
postmean	11.5066	10.5204	14.2818	39.0561	18.5102	8.1112
postsd	2.4759	3.8996	5.5054	7.7194	8.0436	3.1916
lower	6.7871	2.9384	3.6574	24.3880	3.0371	1.9429
upper	16.4632	18.2869	25.1311	54.7345	34.6653	14.5081
Aug 1998 - Mar 2009						
postmean		6.8301		6.1351		
postsd		2.7811		2.6404		
lower		1.4851		1.1631		
upper		12.4653		11.4263		
Apr 2009 - Dec 2023						
postmean	5.8837			8.0133		
postsd	1.7294			3.8458		
lower	2.5758			0.6138		
upper	9.3846			15.7251		
Nov 1971 - Dec 2023						
postmean	6.0283			7.0108	11.0056	3.9490
postsd	1.0366			1.8722	2.2118	0.9831
lower	4.0146			3.4249	6.7792	2.0402
upper	8.0640			10.7145	15.3833	5.8903

The results for the new risk factors are remarkable. In every regime and for each omitted factor, the differences in log-marginal likelihoods are more than 0.69, implying that the model without an intercept is preferred and the selected set of factors in each regime price all of the omitted factors. This gives confidence that our method selects the appropriate set

test of alpha. Finally, the marginal likelihood is a measure of out-of-sample predictive performance of each model, unlike a t-test which measures the extent of departure of an estimator from the null of zero in unseen samples. Due to these differences, frequentist tests can yield different conclusions.

Table 6: **Pricing of the omitted factors by the selected factors in each regime: log-marginal likelihoods of regression models without and with an intercept.** In each regime, the right-hand side variables are the selected risk-factors; the left-hand side variable is the omitted factor displayed in the corresponding column. We report the difference between the regression without an intercept and the regression with an intercept.

	Mkt	SMB	HML	RMW	CMA	MOM
Nov 1971 - Mar 1982						
risk factors	0	1	1	1	1	1
without alpha	180.8431					
with alpha	177.5044					
difference	3.3387					
Apr 1982 - Jul 1998						
risk factors	1	1	1	1	1	1
without alpha						
with alpha						
difference						
Aug 1998 - Mar 2009						
risk factors	0	1	0	1	0	0
Without alpha	198.9939		223.0305		256.5941	156.1655
With alpha	197.3436		221.0403		254.7121	154.6532
Difference	1.6503		1.9902		1.8821	1.5123
Apr 2009 - Dec 2023						
risk factors	1	0	0	1	0	0
without alpha		350.2882	287.9503		353.9200	157.9016
with alpha		348.6746	286.7828		353.0529	156.5496
difference		1.6136	1.1675		0.8671	1.3520

of risk factors in each regime.

5.2. Time-varying factor allocations in maximum Sharpe ratio portfolio

The weights of the maximum Sharpe ratio portfolio that is constructed from the selected set of risk factors are shown in Figure 8. The solid blue line shows the time-varying estimated weights from the breakpoint model and the dashed red line gives time-invariant weights from the no-break model. We find clear evidence of time-variation in the weights allocated across the six factors in the maximum Sharpe ratio which is concealed when breaks are precluded.

We see that the optimal weighting of the market and profitability factors have increased over time such that they are about 40 and 60 percent now. The remaining four factors have generally experienced declines from about 20 percent at the start of the period to zero now, although size received a large 50 percent allocation in the third regime. The no-break model allocated about 40 percent of the weight to CMA with the remaining 60 percent split about

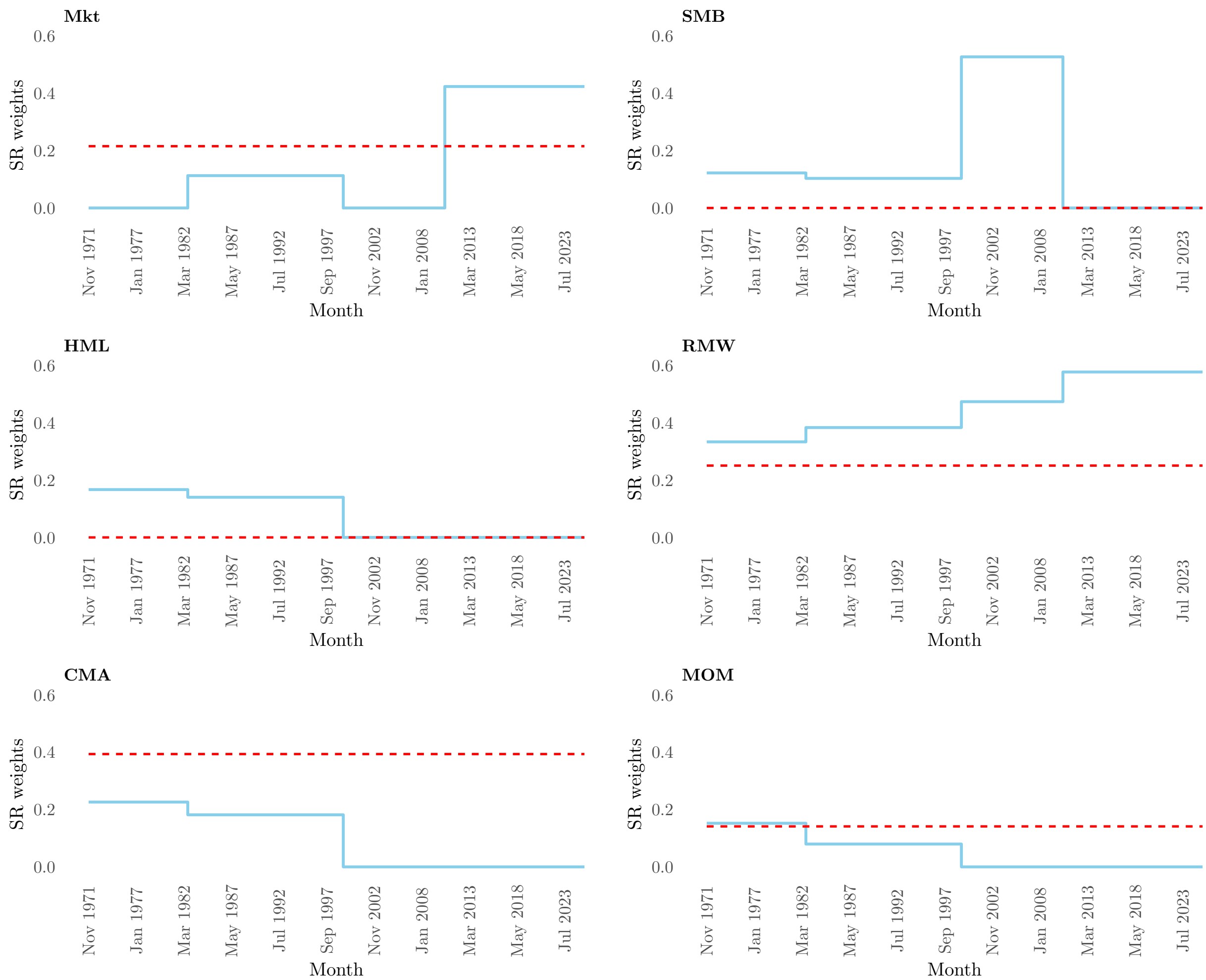

Figure 8: Each window of this Figure displays the time-varying (solid blue line) and time-invariant (dashed red line) estimated weights for the corresponding factor labeled in the subtitle in the maximum Sharpe ratio portfolio constructed from the selected factors. The time-varying estimates are from the selected model in each regime and the time-invariant estimates are from the same model that precludes breaks.

equally between MKT, RMW, and MOM. These findings have important implications for investment strategies.

6. Conclusion

An extensive literature has proposed a multitude of factors that claim to price the cross section of expected returns. This proliferation of factors has led to a more recent literature that attempts to impose discipline on these factors in various ways and hence tame the 'factor zoo'. This paper operationalizes a simple yet novel point, overlooked in the literature, that it is important to account for occasional infrequent shifts or breaks in the set of risk factors. This is because using all available historical data tends to detect factors that were once relevant but are no longer, thereby overstating the relevant set of factors.

Since the date on which the risk factor set changes is unknown, it must be estimated. Existing breakpoint methods are not suitable for this setting, since they only allow the parameters of the model to change, but assume the model itself remains the same over time. Similarly, existing model selection methods do not account for changes in the model over time. We develop the first formal estimation procedure for this setting (either Bayesian or frequentist) that performs an exhaustive search to identify the optimal risk factor collection that is allowed to change at multiple unknown times.

Empirically, we find evidence of three breaks in a six-factor model (Fama-French 5-factors plus momentum) since 1963 that occur in 1982, 1998, and 2009. Our break dates correspond approximately to the end of the "monetarist policy experiment" implemented by Paul Volcker, the surge of the Internet revolution, and digitization of financial markets, and the Global Financial Crisis, suggesting that the optimal set of risk factors can undergo major changes in the presence of such events. We document clear evidence of a shift towards sparse/parsimonious factor models in the post-1998 period. Since the GFC, only two factors (MKT and RMW) have been preferred.

The approach that precludes breaks and that performs factor selection using all the available data spuriously detects an additional two factors (MOM and CMA) that were only relevant until 1998. Our findings have clear implications for the 'factor zoo' literature: those who do not use only the most recent data when conducting factor selection will spuriously detect additional factors that are no longer relevant. This offers one partial explanation for the 'factor zoo': too many factors are being selected because they are fitting pre-break data. In addition, sparse/parsimonious models appear to be favorable for investors who wish to build investment strategies based on such factors today.

Our framework should open new avenues for future research on the challenging problem of detecting risk factors. A R package to implement the methodology is available from the home page of the first author.

References

ADRIAN, T., CRUMP, R. K. and MOENCH, E. (2015). Regression-based estimation of dynamic asset pricing models. *Journal of Financial Economics*, **118** (2), 211–244.

ANG, A. and KRISTENSEN, D. (2012). Testing conditional factor models. *Journal of Financial Economics*, **106** (1), 132–156.

BEKAERT, G., HARVEY, C. R. and LUMSDAINE, R. L. (2002). Dating the integration of world equity markets. *Journal of Financial Economics*, **65** (2), 203–247.

BESSEMBINDER, H., BURT, A. and HRDLICKA, C. M. (2021). Time series variation in the factor zoo. *SSRN Working Paper*.

BIANCHI, D., BÜCHNER, M. and TAMONI, A. (2019). What matters when? time-varying sparsity in expected returns. *Time-Varying Sparsity in Expected Returns (August 17, 2019). WBS Finance Group Research Paper*.

CHIB, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, **90** (432), 1313–1321.

— (1998). Estimation and comparison of multiple change point models. *Journal of Econometrics*, **86**, 221–241.

— (2024). Model uncertainty and change points. *Working Paper*.

—, SHIN, M. and SIMONI, A. (2018). Bayesian estimation and comparison of moment condition models. *Journal of the American Statistical Association*, **113** (524), 1656–1668.

— and ZENG, X. (2020). Which factors are risk factors in asset pricing? A model scan framework. *Journal of Business & Economic Statistics*, **38**, 771–783.

—, — and ZHAO, L. (2020). On comparing asset pricing models. *Journal of Finance*, **75**, 551–577.

—, ZHAO, L. and ZHOU, G. (2024). Winners from winners: A tale of risk factors. *Management Science*, **70** (1), 396–414.

CHORDIA, T., ROLL, R. and SUBRAHMANYAM, A. (2011). Recent trends in trading activity and market quality. *Journal of Financial Economics*, **101** (2), 243–263.

COCHRANE, J. H. (2011). Presidential address: Discount rates. *Journal of Finance*, **66** (4), 1047–1108.

Fama, E. F. and French, K. R. (2018). Choosing factors. *Journal of Financial Economics*, **128** (2), 234–252.

— and — (2021). The value premium. *Review of Asset Pricing Studies*, **11** (1), 105–121.

Feng, G., Giglio, S. and Xiu, D. (2020). Taming the factor zoo: A test of new factors. *Journal of Finance*, **75** (3), 1327–1370.

Ferson, W. E. and Harvey, C. R. (1991). The variation of economic risk premiums. *Journal of Political Economy*, **99** (2), 385–415.

Freyberger, J., Neuhierl, A. and Weber, M. (2020). Dissecting characteristics non-parametrically. *Review of Financial Studies*, **33** (5), 2326–2377.

Gagliardini, P., Ossola, E. and Scaillet, O. (2016). Time-varying risk premium in large cross-sectional equity data sets. *Econometrica*, **84** (3), 985–1046.

Green, J., Hand, J. R. M. and Zhang, X. F. (2017). The characteristics that provide independent information about average u.s. monthly stock returns. *Review of Financial Studies*, **30** (12), 4389–4436.

Griffin, J. M., Harris, J. H., Shu, T. and Topaloglu, S. (2011). Who drove and burst the tech bubble? *Journal of Finance*, **66** (4), 1251–1290.

Gu, S., Kelly, B. and Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, **33** (5), 2223–2273.

Harvey, C. R., Liu, Y. and Zhu, H. (2016). ... and the cross-section of expected returns. *Review of Financial Studies*, **29** (1), 5–68.

Hou, K., Xue, C. and Zhang, L. (2020). Replicating anomalies. *Review of financial studies*, **33** (5), 2019–2133.

Kelly, B. T., Pruitt, S. and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, **134** (3), 501–524.

Lettau, M. and Van Nieuwerburgh, S. (2008). Reconciling the return predictability evidence. *Review of Financial Studies*, **21** (4), 1607–1652.

Lo, A. W. (2004). The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management*, **30th Anniversary Issue**, 15–29.

McLean, R. D. and Pontiff, J. (2016). Does academic research destroy stock return predictability? *Journal of Finance*, **71** (1), 5–32.

Pástor, L. and Stambaugh, R. F. (1999). Costs of equity capital and model mispricing. *The Journal of Finance*, **54** (1), 67–121.

— and Stambaugh, R. F. (2001). The equity premium and structural breaks. *Journal of Finance*, **56** (4), 1207–1239.

Rapach, D. E., Strauss, J. K. and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, **23** (2), 821–862.

Schwert, G. W. (2003). Anomalies and market efficiency. *Handbook of the Economics of Finance*, **1**, 939–974.

Smith, S. C. and Timmermann, A. (2021). Break risk. *The Review of Financial Studies*, **34** (4), 2045–2100.

— and — (2022). Have risk premia vanished? *Journal of Financial Economics*, **145** (2), 553–576.

Viceira, L. M. (1997). Testing for structural change in the predictability of asset returns. *Manuscript, Harvard University*, **4** (3.5), 3–0.