Finance and Economics Discussion Series

Federal Reserve Board, Washington, D.C.
ISSN 1936-2854 (Print)
ISSN 2767-3898 (Online)

Factor Selection and Structural Breaks

Siddhartha Chib and Simon C. Smith

2024-037

Please cite this paper as:
Chib, Siddhartha, and Simon C. Smith (2024). "Factor Selection and Structural Breaks," Finance and Economics Discussion Series 2024-037. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2024.037.

Factor Selection and Structural Breaks

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

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

Draft: May 31, 2024

Abstract

We develop a new approach to select risk factors in an asset pricing model that allows the set to change at multiple unknown break dates. Using the six factors displayed in Table 1 since 1963, we document a marked shift towards parsimonious models in the last two decades. Prior to 2005, five or six factors are selected, but just two are selected thereafter. This finding offers a simple implication for the factor zoo literature: ignoring breaks detects additional factors that are no longer relevant. Moreover, all omitted factors are priced by the selected factors in every regime. Finally, the selected factors outperform popular factor models as an investment strategy.

Keywords: Model comparison, Factor models, Structural breaks, Anomaly, Bayesian analysis, Discount factor, Portfolio analysis, Sparsity.

JEL classifications: G12, C11, C12, C52, C58

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

We thank Daniele Bianchi and Andy Neuhierl. Any remaining errors are our own. The views expressed in this paper are those of the authors and do not necessarily reflect the views and policies of the Board of Governors or the Federal Reserve System. Center for Research in Security Prices data were obtained by Siddhartha Chib under the purview of Washington University licenses.

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 huge 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).

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.[2] Otherwise, 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;

[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.'

[2] For example, the publication effect of Schwert (2003), and/or the adaptive efficient market hypothesis of Lo (2004), may cause the set of risk factors to change. The set of risk factors may also change due, for example, to the technological revolution in financial markets towards the end of the twentieth century, shifting monetary policy regimes that led to the anchoring of inflation expectations, or regulatory changes.

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 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.[3] 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.[4] 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).

Our methodology generalizes the framework of Chib and Zeng (2020) – who developed a Bayesian model selection approach for time-invariant factor selection – by blending it with the Bayesian breakpoint approach in the context of model uncertainty developed by Chib (2024), producing a single unified framework which estimates the selected risk factors and allows this selected set to change at multiple unknown break dates. Note that a Bayesian approach is well suited to this problem because it can allow for both abrupt and gradual changes, depending on the uncertainty surrounding the break date. A Bayesian approach

[3]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.

[4]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).

also inherently protects against problems associated with multiple tests (Kozak *et al.* 2020; Jensen *et al.* 2023; Bryzgalova *et al.* 2023). We perform an exhaustive search across *all possible* asset pricing models implied by the starting set of risk factors and all possible break dates for a given number of breaks, identifying the optimal subset of potential factors that can price most (if not all) of the remaining factors in each regime.[5] Our exhaustive search circumvents the risk of getting stuck at local maximia that is associated with stochastic search algorithms.[6]

In our empirical analysis, we focus on the six-factor model of Fama and French (2018).[7] Using monthly data from July 1963 through December 2023, our method identifies three breaks corresponding to a regime lasting for 15 years on average.[8] The breaks occur in March 1975, October 1995, and September 2005.[9]

The set of risk factors changes after each of these breaks. At least five factors are selected in the first three regimes (up to 2005), while only two factors (market and profitability) are selected in the final regime (post-2005).[10] 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.[11]

[5]Our method also performs inference over the number of breaks.

[6]While the conventional approach to test the pricing ability of risk-factors is to use various test assets or portfolios, following Chib and Zeng (2020) we leverage the intuition that if a subset of the available factors are found to be risk factors, then those factors, by virtue of being risk factors, should price the complementary set of non risk factors.

[7]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.

[8]We consider other numbers of breaks, but find three to be optimal.

[9]The break in 1975 corresponds to the oil price shocks of the 1970s and the corresponding high inflationary period that was only stopped when a sharp contractionary monetary policy regime was subsequently implemented. October 1995 coincides with the Internet revolution and the tech boom on the NASDAQ (Griffin *et al.* 2011). This break also coincides with a period of dramatic changes in market efficiency that has been documented by Chordia *et al.* (2011). The September 2005 break corresponds to a little before the onset of the Global Financial Crisis.

[10]All but the size factor are selected in the first regime (1963-1975), all six factors are selected in the second (1975-1995), and all but the value factor are selected in the third (1995-2005).

[11]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

Moreover, the median number of factors selected in the best performing ten models in the three regimes up to 2005 is five, but this falls to 2.5 in the final regime. This clearly indicates a shift to more parsimonious models in the most recent two decades.[12] In fact, the Capital Asset Pricing Model (CAPM) – which performs poorly up to 2005 – is in the top ten models after 2005 and outperforms the 3- and 5-factor models of Fama-French (and both of those models plus momentum).

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. Post-2005, constructing the tangency portfolio that consists of the selected factors and the individual stocks that are not priced by those factors generates a Sharpe ratio of 2.74. This is much higher than the corresponding Sharpe ratios (which range from 0.87 to 1.82) generated from the 3- and 5-factor models of Fama-French, the same two models plus momentum, and the CAPM. The two risk factors that our procedure has isolated since 2005 – market and profitability – capture important systematic risks. The role of the market factor as a systematic risk factor is arguably unquestioned. The profitability factor captures the part of the cross section of expected returns that covaries with profitability. 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.[13] 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 is not

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.

[12] Kelly *et al.* (2019) use Instrumented Principal Components Analysis to document that just five latent factors can outperform existing factor models.

[13] 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).

only dependent on the magnitude of the associated risk premia, but also on the stability of their risk premia over time.[14] We find clear time-variation in the risk premia for all six factors since 1963. For example, the value premium was 5.6% from 1963 to 1975 but decreased to 4.3% from 1975 to 1995 (Fama and French 2021). Since 1995, the value factor has not been selected as a risk factor. The implied weights on the value factor in the maximum Sharpe ratio portfolio therefore declined from 18 percent (1963-1975) to 15 percent (1975-1995) and have been zero since. This indicates 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. We present the first approach (either Bayesian or frequentist) to do so.[15]

The rest of the paper is organized as follows. In Section 2 we detail our methodology. In Section 3 we present evidence of breaks and the regime-specific selected factors and their risk premia estimates. Section 4 has the pricing performance and investment implications of our selected factor collection, and Section 5 concludes.

[14] 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 portfolios, while others may pursue very large payoffs. Some investors may pursue "buy-and-hold" strategies, and others may periodically rebalance to target certain weights.

[15] Bianchi *et al.* (2019) also document evidence of time-varying sparsity in factor models.

2. Methodology

We now set out 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 our prior specification.

2.1. Economic Sources of Breaks in the Factor Model

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

$$M_t = 1 - b'(f_t - \lambda)$$

where b is the vector of market prices of factor risks and λ 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

$$M_t = 1 - b_t'(f_t - \lambda_t)$$

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 b_t would be zero and the corresponding elements of f_t would drop out of the SDF, i.e., cease to be risk factors.

To describe this situation, let $x_t \subseteq f_t$ denote a subset of f_t with non-zero market prices of factor risks. Suppose that the market prices 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

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

The objectives of the analysis are to find

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

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 which such decomposition is the best.[16] The risk factors in that best decomposition are then taken to be the risk factors best supported by the data.

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

$$\begin{aligned} x_{j,t} &= \lambda_j + u_{j,t} \\ y_{j,t} &= \Gamma_j x_{j,t} + \varepsilon_{j,t} \ , \qquad t = 1, \ldots, T \ , \end{aligned} \tag{3}$$

where the errors are distributed as multivariate Gaussian

$$u_{j,t} \sim N(0, \Omega_j) \ , \qquad \varepsilon_{j,t} \sim N(0, \Sigma_j). \tag{4}$$

Let the unknown parameters in this model be denoted by

$$\boldsymbol{\theta}_j = (\boldsymbol{\lambda}_j, \Omega_j, \Gamma_j, \Sigma_j) \, . \tag{5}$$

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 $\Omega_j, \Gamma_j, \Sigma_j$ are derived by change-of-variable from a *single* inverse Wishart prior placed on the matrix Ω_j in the model where all factors

[16] Marginal likelihoods are Bayesian objects that are calculated by integrating out the parameters from the sampling density with respect to the prior of the parameters.

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, which means that it is 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 $\mathbb{M}_j$ is given by

$$\text{marglik}(f|\mathbb{M}_j) = \int N(x_j|\lambda_j, \Omega_j) N(y_j|\Gamma_j x_j, \Sigma_j) d\pi(\boldsymbol{\theta}_j), \qquad j \leq J. \tag{6}$$

These are 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. This break occurs at an unknown location t_1^* that separates the sample data into regimes $s \in \{1, 2\}$. A set of risk factors (x_t^1) enters the SDF in the first regime (from time periods $t = 1, \ldots, t_1^*$) and another set (x_t^2) enters in the second regime (from time periods $t = t_1^* + 1, \ldots, T$).[17] The objective is to estimate the timing of the break (t_1^*) and the identities of the risk factors in the first regime (x_t^1) and the second (x_t^2) regime.

To infer the break date, we focus on the quantity

$$\text{marglik}(f_{1,t_1^*}, f_{t_1^*+1,T}|t_1^*) \tag{7}$$

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 choose the break date with the largest value of this marginal likelihood.

[17]The risk factor set is stable within each regime.

The problem in calculating the preceding quantity is that we do not have the data-generating process (DGP) on either side of the split. In other words, 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 f_t into x_t and y_t, on either side of t_1^*. On the left, we denote the models by $\mathbb{M}_{j,1}$ and on the right by $\mathbb{M}_{k,1}$, for $(j,k) = 1, ..., J = 2^K - 1$. When $j = k$ the splits are identical but the parameters of the model are different. Just as we did in Equation (3), the jth model in regime s, $s = 1, 2$ takes the form

$$
\begin{aligned}
x_{j,t,s} &= \lambda_{j,s} + u_{j,t,s} \\
y_{j,t,s} &= \Gamma_{j,s} x_{j,t,s} + \varepsilon_{j,t,s} \\
u_{j,t,s} &\sim N(0, \Omega_{j,s}) \\
\varepsilon_{j,t,s} &\sim N(0, \Sigma_{j,s}), \qquad t \in T_{s,1}\ , \qquad s = 1, 2,
\end{aligned}
\tag{8}
$$

where $T_{1,1} = (1, 2, ..., t_1^*)$ and $T_{2,1} = (t_1^* + 1, ..., T)$. We denote the unknown parameters in these models by $\boldsymbol{\theta}_{j,s} = (\boldsymbol{\lambda}_{j.s}, \Omega_{j,s}, \Gamma_{j,s}, \Sigma_{j,s})$. Note that each of these models has a different set of risk factors and a distinct 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 $\mathbb{M}_{j,s}$ is given by

$$
\begin{aligned}
&\text{marglik}(f_{s,m} | \mathbb{M}_{j,s}, t_1^*) \\
&= \int N(x_{j,t,s} | \lambda_{j,s}, \Omega_{j,s}) N(y_{j,t,s} | \Gamma_{j,s} x_{j,t,s}, \Sigma_{j,s}) d\pi(\boldsymbol{\theta}_{j,s})\ , \qquad j \leq J,\ s = 1, 2
\end{aligned}
\tag{9}
$$

which we calculate by the method of Chib (1995a).

Now by extending the argument and marginalization the marginal likelihood in Equation

(7) can be written as

$$\text{marglik}(f_{1,t_1^*}, f_{t_1^*+1,T}|t_1^*) = \sum_{k=1}^{J}\sum_{j=1}^{J}\text{marglik}(f_{1,t_1^*}, f_{t_1^*+1,T}|\mathbb{M}_{j,1}, \mathbb{M}_{k,1}, t_1^*)\,\Pr(\mathbb{M}_{j,1})\,\Pr(\mathbb{M}_{k,1}) \tag{10}$$

$$= \frac{1}{J^2}\sum_{k=1}^{J}\sum_{j=1}^{J}\text{marglik}(f_{1,t_1^*}|\mathbb{M}_{j,1}, t_1^*)\text{marglik}(f_{t_1^*+1,T}|\mathbb{M}_{k,2}, t_1^*) \tag{11}$$

where in the second line we have assumed equal prior probabilities of models and the fact that the joint 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:

$$\text{marglik}(f_{1,m}, ..., f_{m+1,m}|t_1, \ldots, t_m). \tag{12}$$

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

$$f_{s,m} = \{f_t : t_{s-1} < t \leq t_s\}, \qquad s = 1, ..., m + 1. \tag{14}$$

Once again, we consider all possible splits of f_t into x_t and 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

$$\begin{aligned}
x_{j,t,s} &= \lambda_{j,s} + u_{j,t,s} \\
y_{j,t,s} &= \Gamma_{j,s} x_{j,t,s} + \varepsilon_{j,t,s} \\
u_{j,t,s} &\sim N(0, \Omega_{j,s}) \\
\varepsilon_{j,t,s} &\sim N(0, \Sigma_{j,s}), \qquad t \in T_{s,m}.
\end{aligned} \tag{15}$$

Denoting the unknown parameters in these models by $\boldsymbol{\theta}_{j,s} = (\boldsymbol{\lambda}_{j,s}, \Omega_{j,s}, \Gamma_{j,s}, \Sigma_{j,s})$, the marginal likelihood of $\mathbb{M}_{j,s}$ is given by

$$\begin{aligned}
&\text{marglik}(f_{s,m}|\mathbb{M}_{j,s}, t_1^*, \ldots, t_m^*) \\
&= \int N(x_{j,t,s}|\lambda_{j,s}, \Omega_{j,s}) N(y_{j,t,s}|\Gamma_{j,s} x_{j,t,s}, \Sigma_{j,s}) d\pi\left(\boldsymbol{\theta}_{j,s}\right) \ , \qquad j \leq J, \ s = 1, \ldots, m + 1
\end{aligned} \tag{16}$$

which we calculate by the method of Chib (1995a).

The next step is to calculate the marginal likelihood of all the data for given pairings of models from each of the $m + 1$ regimes. There are $J^{(m+1)}$ such pairings in all regimes. The marginal likelihood in Equation (12) can be written as

$$\begin{aligned}
&\text{marglik}(f_{1,m}, ..., f_{m+1,m}|\mathbb{M}_{j_1,1}, \mathbb{M}_{j_2,1}, ..., \mathbb{M}_{j_J,1}, ..., \mathbb{M}_{j_1,m+1}, \mathbb{M}_{j_2,m+1} ..., \mathbb{M}_{j_J,m+1}, t_1, \ldots, t_m) \\
&= \prod_{s=1}^{m+1} \text{marglik}(f_{s,m}|\mathbb{M}_{j_s,s}, t_1, \ldots, t_m). \qquad (17)
\end{aligned}$$

We can get the desired marginal likelihood by summing the right hand side over all possible pairings of models. If $m = 3$ and $J = 63$, as in one of our cases we consider, there are more than 15 million such model combinations. Thus,

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

This is the marginal likelihood for the break dates $t_1, \ldots, t_m$.[18] The calculation is repeated for all possible locations of the m breaks and all possible combinations of the J models across the corresponding $m + 1$ regimes. 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.[19] Since our Bayesian approach is based on the marginal likelihood which penalizes overparameterization, it will inherently guard against overfitting the number of breaks.[20]

[18]Again, we have assumed equal prior probabilities of models and the fact that the joint factors into independent components given the models. 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.

[19]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.

[20]The identity in Chib (1995b) 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 param-

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. We now set out the prior for the no break case. The same prior is constructed for the parameters in each regime of the breakpoint model.

Formally, models are defined by the parameters

$$\boldsymbol{\theta}_j = (\boldsymbol{\lambda}_j, \boldsymbol{\psi}_j)$$

where

$$\boldsymbol{\psi}_j = \begin{cases} \boldsymbol{\Omega}_1 & j = 1 \\ (\boldsymbol{\Omega}_j, \boldsymbol{\Gamma}_j, \boldsymbol{\Sigma}_j) & j \geq 2. \end{cases}$$

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

$$\boldsymbol{\psi}_1 = \boldsymbol{\Omega}_1,$$

is exactly equal to the number of free parameters in

$$\boldsymbol{\psi}_j \quad j \geq 2$$

and that there is a one-to-one mapping between them. We specify a prior on $\boldsymbol{\Omega}_1$ and from this single prior, by change of variable, we derive the prior for $\boldsymbol{\psi}_j$, $j \geq 2$. In this way, the priors across models are equalized.

Finally, the prior of $\boldsymbol{\lambda}_j$ is found from a training sample approach. For further details, see Chib and Zeng (2020).

---

eters. Thus, more complex models will have a bigger penalty.

3. 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.

3.1. Data

Our analysis focuses on the six factors of Fama and French (2018).[21] We use monthly data from July 1963 through December 2023, available from Ken French's website.

3.2. 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 fashion, which is hard 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 1 displays whether the corresponding factor labeled in the subcaption is selected (the indicator variable equals one) or omitted at a given time point.

[21] Details of these factors are provided in Table 1.

These selections are obtained from estimating recursively our methodology that excludes breaks but uses a fixed rolling window length of five years. We see that factors are frequently omitted and then selected again shortly after.

Figure 2 displays the corresponding total number of factors selected at each time point from this rolling-window approach. We see that the number of factors selected in the model varies between one and five and is changing very frequently. Such frequent changes in the sparsity of the factor model 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.

3.3. Evidence of breaks

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 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 1975, October 1995, and September 2005. These break dates correspond closely to major events such as 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 the final break occurs just before the onset of the Global Financial Crisis. The post-2005 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.

3.3.1. 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 it captures uncertainty surrounding the break dates. The top panel of Figure 3 displays the posterior probability – measured in percent – that the first break identified by our model occurs in a given month. We see that while most of the probability (about 50%) is assigned to the optimal break date of March 1975, there is uncertainty around this, with the remainder of the probability being distributed throughout the year. The lower panel displays the corresponding cumulative posterior probability, measured in percent, that our model has identified no breaks (dotted line) or one break (solid line). As of December 1974, one is fully confident that we are still in the first regime, as zero posterior probability is assigned to one break. However, with each subsequent month the probability assigned to one break increases until, by March 1975, this scenario receives more weight than the no-break case. By August, one is almost certain that we have transitioned to the second regime.

This is a particularly attractive feature of our approach, because our methodology is flexible enough to allow for both instant shifts and gradual changes. The extreme case of no uncertainty surrounding the break date estimate implies an abrupt shift. As this uncertainty increases, the shift becomes more and more smooth. Popular approaches to capture such changes can typically accommodate one of these scenarios, but not both. For instance, a time-varying parameter model captures gradual changes, while a frequentist break-point approach captures abrupt changes.

The uncertainty surrounding the break date estimate can be incorporated into risk premia and price of risk estimates from the factor model, and thus our framework can allow for both abrupt and gradual changes in risk premia and price of risk estimates.

3.4. 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 (1963-1975), the optimal model is the one which includes all six potential factors except the size factor. This collection of risk factors has a log marginal likelihood of 526.451. Each of the top five models in this regime omits the size factor, providing 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 market factor and the momentum factor. The upper left window of Figure 4 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 SMB, is shown in blue. Other models of interest are colored red. We see that dense models that include more factors (FF6 and FF5) tend to perform better than sparse models (FF3 and CAPM). Note that the FF6 model is the sixth best model out of 63.

The upper left window of Figure 5 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 is colored blue. The remaining 62 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). The circles plotted in the figure therefore move from the top-left towards the bottom-right with the median number of factors selected in the top ten performing models being equal to four. This finding suggests that dense factor models are preferred to sparse models during this period.

In the second regime (1975-1995), the FF6 model is the best performing model with a log marginal likelihood of 3586.089. Each of the five best-performing models includes at least four factors, and usually at least five. The HML and RMW factors are the only ones selected in each of the five best models. Once again, popular dense models tend to perform well (FF6 and FF5), while popular sparse models (FF3 and CAPM) 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 5). The median number of factors in the top ten performing models is equal to five, even higher than in the first regime.

The third regime (1995-2005) displays the same dynamics as the first two. A five-factor model that omits HML is preferred with a log marginal likelihood equal to 1384.963. The FF6 model is the second-best model, and each of the top five models contains at least five factors. The MKT, SMB, and RMW factors are selected in each of the top five models. Once again, popular dense models (FF6 and FF5) outperform popular sparse models (FF3 and CAPM). The pattern in which dense models tend to outperform sparse models continues to persist in this regime, with the median number of factors selected in the top ten models once again equal to five.

However, in the final regime (2005-2023), the pattern changes markedly. The best-performing model now selects only two factors: market and profitability. The top five models contain between two and at most four factors. Market and profitability are selected in each of these five models, providing strong evidence that these factors are relevant, while SMB is not selected in any of them. It is striking that in this regime the simple CAPM is one of the best performing models, above denser models such as FF3, FF5 and FF6. Across all the possible 63 models in this regime, we no longer see any evidence that dense models outperform sparse models. If anything, the pattern reverses, with the median number of factors in the top ten performing models equal to just 2.5. A model selection approach that ignores breaks and uses all available historical data (1963-2023) identifies a four-factor model (MKT, RMW, CMA, MOM) as the best performing model with a log marginal likelihood of 8022.842 (top row of the bottom panel in Table 3). Here we see that popular dense models (FF6 and FF5) outperform popular sparse models (CAPM and FF3). Across all 63 possible models, we see a clear pattern in which dense models tend to outperform sparse models, as observed in the first three regimes using the model that allows for breaks. The median number of factors selected in the top ten models is four.

These findings make a simple, yet novel point. Until 2005, dense models were preferred, but since 2005 there has been a clear shift towards parsimony with sparse models performing better. Ignoring breaks conceals this changing dynamic and would lead one to spuriously believe that dense models are still preferable 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 post-2005 data in their future analyses.

3.5. 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.

Each window of Figure 6 displays the estimated posterior mean of the risk premia (expressed in annualized percent) for the corresponding factor labeled in the caption. Parts of the solid black line that are "missing" correspond to that factor not being selected in the given regime. Similarly, the dashed black line is omitted for HML and SMB altogether since they are not selected in the model that precludes breaks.

We see that the equity premium (top-left window) is estimated to be about 7 percent 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 highest value of around ten percent in the most recent regime (post-2005).

The only other factor that is selected throughout the sample and in the current regime is the profitability factor (middle-right window). We see that the estimated premium on this factor has undergone time variation, increasing from a low value to a value of about 4% in 1975, where it has remained roughly ever since. The no-break model smooths through all these regimes, estimating a premium of around 3.5 percent throughout the sample.

We also document clear evidence of a decrease in value premium (middle left) throughout our sample, corroborating evidence documented in earlier studies (Fama and French 2021). Specifically, we see that it falls from close to 6 percent until 1975 to just above 4 percent from 1975 until 2005. Since 2005, the value factor has been omitted altogether.

Both the investment and the momentum factor risk premia estimated from the breakpoint model reveal larger premia than those from the time-invariant model. By pooling information across this final regime, the model that excludes breaks estimates a lower premia in the earlier periods than the break model.

The size factor is omitted in the first and final regimes and is estimated to have increased across the middle two regimes from about two percent (1975-1995) to about six percent (1995-2005). The time-invariant model omits the factor altogether, ruling out the possibility of any comparison. In short, we find evidence of time-variation in all six-factor risk premia.

The four top panels of Table 4 show, for each of the four regimes, a range of risk premia estimates for the factors selected by the optimal model in that regime. Specifically, we report the posterior mean, standard deviation, and median of the risk premia estimates. We also report the lower and upper estimates that correspond to the 95 percentiles of the posterior distribution. The bottom panel displays the same information for the model that precludes breaks. Figure 7 displays the corresponding estimated posterior densities of the risk premia.

In the most recent regime, we see that the only two factors that are selected – market and profitability – both generate significant risk premia in the formal sense. Both are also significant in the second regime (1975-1995).[22]

Turning to the market prices of factor risks – which are the weights of the risk factors in the SDF – Figure 8 also shows clear evidence of the time variation in the market price of risk for the six factors that is hidden when excluding breaks. Once again, Table 5 shows that the market prices of market and the profitability factor risks are significant in the formal sense in the second and third regimes, although the price of risk for the market is also significant in the second regime.

4. Pricing performance and investment implications

4.1. Pricing of the excluded factors

Having revealed the optimal regime-specific risk factor collections, we next explore whether in each regime the selected factors can price the omitted ones. This is not a trivial hurdle because, for example, using the full sample of data, MOM is not priced by the FF-5 factor model; MOM, RMW, CMA are not priced by the FF-3 factor model, and SMB is not priced using the factors selected from the model that precluded breaks (MKT, RMW, CMA, MOM).

To evaluate pricing ability, we fit a sequence of regression models with each excluded factor on the left side and the selected risk factors on the right side. This is performed separately for each regime using only the data from that regime and the factor selection in that regime. For each excluded factor two Bayesian regressions are estimated, one with an

[22]Given that the first and third regimes are relatively short, achieving significance in the formal sense may be challenging due to imprecision in the estimates resulting from the inherent noise in the data.

intercept and one without. The marginal likelihood of each model is then calculated using the method of Chib (1995a). 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, by an application of Jeffreys' rule, we can conclude that the omitted factor is priced by the selected factors in that regime. Formally, surpassing this threshold implies that the posterior odds of the model without an intercept relative to the model with an intercept is at least 2:1.

This Bayesian test is better able to detect the priced factors than any of the frequentist tests in current use, such as those based on the t-stat of the estimated intercept, or average absolute estimated intercept, etc.[23] The results of applying this test to each of the excluded factors are shown in Table 6.

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 of risk factors in each regime.

4.2. Pricing the cross-section and investment implications

We now provide some evidence on the performance of the current set of risk factors (MKT, RMW) in pricing the cross section of stocks in the final regime we identify. Our sample, drawn from the CRSP database[24], consists of 1,992 stocks that have at least 180 months

[23]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 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 fundamental differences, it is possible that frequentist and Bayesian approaches can yield different conclusions.

[24]Source: Center for Research in Security Prices, CRSP 1925 US Stock Database.

of available data during our sample.[25] We isolate the stocks that are not priced by these two risk factors since those are the ones that offer an investment opportunity. We evaluate whether it is possible to form a minimum mean-variance portfolio by combining the selected risk factors with the stocks that they do not price.

For each stock, we apply the Bayesian pricing test described in the previous section. Specifically, for the excess return of each stock i, a pair of Bayesian regression models are estimated, one with an intercept and one without. The marginal likelihood of each model in the pair is computed using the method of Chib (1995a) and the best model in the pair is then selected (the ones with the largest marginal likelihoods). If the log marginal likelihood of the model with an intercept does not exceed that of the model without an intercept by more than 0.69, we conclude that stock i can be priced by the set of risk factors. Otherwise, it cannot be priced and offers an investment opportunity.

All stocks deemed unpriced by this criterion, in conjunction with the corresponding risk factors, are used to form a tangency portfolio, which is also the portfolio with the highest Sharpe-ratio (Sharpe 1994). This highest Sharpe-ratio is given by

$$\text{Sharpe ratio} = \sqrt{\hat{\boldsymbol{\mu}}'\hat{\Omega}^{-1}\hat{\boldsymbol{\mu}}}, \tag{19}$$

where $\hat{\boldsymbol{\mu}}$ and $\hat{\Omega}$ are estimates of the mean and covariance of the given risk-factors and its set of unpriced stocks.

This process is then repeated for a range of popular risk factor collections such as the Fama-French 3- and 5-factor models. The number of such assets, of course, varies by risk factor collection (because the number of risk-factors varies as does the number of stocks priced by different risk-factor collections).

The results of these Sharpe ratios are given in Table 7 for the new collection of risk

[25]Following convention, we drop financial firms, those with share codes beyond 10 and 11, and those with prices below \$5.

factors and the existing collections. The Sharpe ratio for the new risk factors is the highest. We leave further study of this interesting finding to future work because it is not directly related to the main goals of this paper.

Finally, we consider time-variation in the weights of the maximum Sharpe ratio portfolio that is constructed from the selected set of risk factors. These weights are shown in Figure 9. The solid black line shows the time-varying estimated weights and the dashed black line gives time-invariant weights. These weights are expressed as percent, and the corresponding factor is labeled in the subcaption. These weights are also reported in Table 8.

We see that the optimal weighting of the market factor has increased over time from about ten percent in the first half of the sample to around 30 percent in the second half. The model that precludes breaks, however, smooths through all the regimes and allocates a constant 20 percent or so to the market. The model that excludes breaks assigns zero weight to size and value at all times, whereas our approach is more flexible and allocates up to 20 percent to these factors in certain regimes and zero percent in others. The profitability factor receives a substantial weight of close to 30% from the no-breaks model, while our approach reveals that the weight in this factor is as much as 70 percent since 2005. Finally, our approach reduces the weight in the investment and momentum factors to zero since 2005, while the no-break model assigns a constant 40 and 14 percent to these two factors. These findings have important implications for investment strategies.

5. 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 1975, 1995, and 2005. Our break dates correspond approximately to the oil price shocks of the 1970s, the surge of the Internet revolution, and digitization of financial markets, and just before 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-2005 period. Until 2005, the preferred model contained either five or six factors. Since 2005, only two factors (MKT and RMW) have been preferred. Moreover, the median number of factors in the top ten models is four or five in each regime until 2005, but is just 2.5 since 2005.

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 2005. 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.

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*.

Bryzgalova, S., Huang, J. and Julliard, C. (2023). Bayesian solutions for the factor zoo: We just ran two quadrillion models. *Journal of Finance*, **78** (1), 487–557.

Chib, S. (1995a). Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association*, **90** (432), 1313–1321.

— (1995b). 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*.

— 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.

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.

Jensen, T. I., Kelly, B. and Pedersen, L. H. (2023). Is there a replication crisis in finance? *Journal of Finance*, **78** (5), 2465–2518.

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.

Kozak, S., Nagel, S. and Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, **135** (2), 271–292.

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. (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.

Sharpe, W. F. (1994). The sharpe ratio. *Journal of Portfolio Management*, **21** (1), 49–58.

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.

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

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 July 1963 through December 2023. The row displayed in bold font corresponds to the optimal number and timing of breaks.

No. of breaks	Log marg lhood.		Break dates	
0	8022.842			
1	8221.959	May 1998		
2	8258.581	Jan 1981	Jul 1998	
3	**8333.784**	**Mar 1975**	**Oct 1995**	**Sep 2005**
3	8331.619	Jan 1975	Oct 1995	Sep 2005
3	8331.411	Feb 1975	Oct 1995	Sep 2005
3	8332.419	Apr 1975	Oct 1995	Sep 2005
3	8332.126	May 1975	Oct 1995	Sep 2005
3	8330.252	Mar 1975	Aug 1995	Sep 2005
3	8332.951	Mar 1975	Sep 1995	Sep 2005
3	8331.371	Mar 1975	Oct 1995	Oct 2005
3	8329.825	Mar 1975	Oct 1995	Nov 2005

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. lhd.
Jul 1963 - Mar 1975	1	0	1	1	1	1	526.451
	1	0	0	1	1	1	525.692
	1	0	1	0	1	1	524.963
	1	0	1	1	0	1	524.643
	1	0	0	0	1	1	524.594
Apr 1975 - Oct 1995	1	1	1	1	1	1	3586.089
	1	1	1	1	0	1	3584.353
	1	0	1	1	1	1	3583.721
	0	1	1	1	1	0	3583.701
	1	1	1	1	0	1	3582.573
Nov 1995 - Sep 2005	1	1	0	1	1	1	1384.963
	1	1	1	1	1	1	1384.826
	1	1	1	1	0	1	1384.094
	1	1	0	1	1	0	1383.382
	1	1	1	1	1	1	1383.210
Oct 2005- Dec 2023	1	0	0	1	0	0	2848.358
	1	0	0	1	1	0	2848.299
	1	0	0	1	0	1	2847.591
	1	0	1	1	1	0	2847.589
	1	0	0	1	1	1	2847.501
Jul 1963 - Dec 2023	1	0	0	1	1	1	8022.842
	1	0	1	1	1	1	8021.338
	1	0	1	1	0	1	8018.978
	1	1	0	1	1	1	8018.851
	1	0	0	0	1	1	8018.635

Table 4: **Risk premia estimates**. The top four panels of this table display, for each of the four regimes, 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. The bottom panel displays the same information for the model that precludes breaks.

	Mkt	SMB	HML	RMW	CMA	MOM
			Jul 1963 - Mar 1975			
postmean	-0.0008		0.0047	0.0003	0.0035	0.0083
postsd	0.0044		0.0025	0.0017	0.0022	0.0037
lower	-0.0094		-0.0001	-0.0031	-0.0008	0.0009
upper	0.0077		0.0097	0.0036	0.0076	0.0156
			Apr 1975 - Oct 1995			
postmean	0.0067	0.0014	0.0035	0.0035	0.0029	0.0091
postsd	0.0031	0.0017	0.0018	0.0009	0.0012	0.0023
lower	0.0007	-0.0002	-0.0002	0.0002	0.0006	0.0005
upper	0.0127	0.0048	0.0069	0.0055	0.0053	0.0136
			Nov 1995 - Sep 2005			
postmean	0.0047	0.0049		0.0036	0.0052	0.0092
postsd	0.0049	0.0042		0.0044	0.0029	0.0063
lower	-0.0051	-0.0034		-0.0048	-0.0007	-0.0032
upper	0.0143	0.0132		0.0124	0.0109	0.0216
			Oct 2005 - Dec 2023			
postmean	0.0089			0.0033		
postsd	0.0035			0.0015		
lower	0.0020			0.0004		
upper	0.0158			0.0062		
			Jul 1963 - Dec 2023			
postmean	0.0059			0.0029	0.0029	0.0058
postsd	0.0019			0.0009	0.0008	0.0018
lower	0.0022			0.0014	0.0014	0.0023
upper	0.0095			0.0046	0.0046	0.0093

Table 5: **Price of risk estimates**. The top four panels of this table display, for each of the four regimes, 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. The bottom panel displays the same information for the model that precludes breaks.

	Mkt	SMB	HML	RMW	CMA	MOM
			Jul 1963 - Mar 1975			
postmean	5.194		8.794	16.154	12.114	6.857
postsd	2.664		5.419	7.720	7.397	2.632
lower	-1.478		-5.027	-2.949	-6.579	0.293
upper	12.289		23.404	37.558	32.272	14.046
			Apr 1975 - Oct 1995			
postmean	7.427	6.848	13.8115	35.669	19.504	6.190
postsd	2.229	3.519	5.478	7.274	8.178	2.638
lower	1.845	-1.969	-0.223	17.794	-0.722	-0.474
upper	13.308	16.315	28.214	55.325	41.588	13.145
			Nov 1995 - Sep 2005			
postmean	10.076	6.044		8.758	10.508	3.152
postsd	3.271	3.341		3.819	4.605	1.855
lower	2.232	-2.300		-0.649	-0.963	-1.595
upper	19.004	15.065		19.054	22.859	8.052
			Oct 2005 - Dec 2023			
postmean	4.475			9.855		
postsd	1.591			3.859		
lower	0.529			0.289		
upper	8.690			20.215		
			Jul 1963 - Dec 2023			
postmean	5.935			6.769	11.109	3.815
postsd	1.036			1.873	2.228	0.984
lower	3.307			1.957	5.458	1.311
upper	8.701			11.603	16.947	6.332

Table 6: **Pricing of the omitted factors by the selected factors in each regime: log-marginal likelihoods of regression models with and without 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. The regressions displayed in the first row contain an intercept, while those in the second do not. We report the difference between the regression without an intercept and the regression with an intercept.

	SMB	HML	CMA	MOM
Jul 1963 - Mar 1975				
with α	256.1637			
without α	257.3031			
difference	1.139			
Apr 1975 - Oct 1995				
with α				
without α				
difference				
Nov 1995 - Sep 2005				
with α		257.853		
without α		260.123		
difference		2.269		
Oct 2005- Dec 2023				
with α	461.053	385.639	490.994	334.839
without α	462.639	386.851	492.429	336.083
difference	1.586	1.211	1.436	1.243

Table 7: **Sharpe ratio of tangency portfolio.** Sharpe-ratios of the tangency portfolio for assets composed of the selected risk-factors in the final regime and the stocks (out of 1,992) that are not priced by those risk-factors in the corresponding regime. The results show the portfolio consisting of the selected risk-factors, plus its unpriced stocks, has the highest Sharpe-ratio across the different sets of risk-factors.

Risk factors	Sharpe ratio
6-factor model	1.69
5-factor model	1.37
4-factor model	0.87
3-factor model	1.82
CAPM	1.73
MKT+RMW	2.74

Table 8: **Factor weights in maximum Sharpe ratio portfolio**. The top four rows of this table display the weight (expressed in percent) allocated to each of the selected factors in the optimal model chosen in each of the four regimes. The bottom row displays the same information for the model that precludes breaks.

	Mkt	SMB	HML	RMW	CMA	MOM
Jul 1963 - Mar 1975	10.58		17.91	32.89	24.67	13.96
Apr 1975 - Oct 1995	8.30	7.66	15.44	39.88	21.80	6.92
Nov 1995 - Sep 2005	26.15	15.68		22.73	27.27	8.18
Oct 2005- Dec 2023	31.23			68.77		
Jul 1963 - Dec 2023	21.48			24.50	40.21	13.81

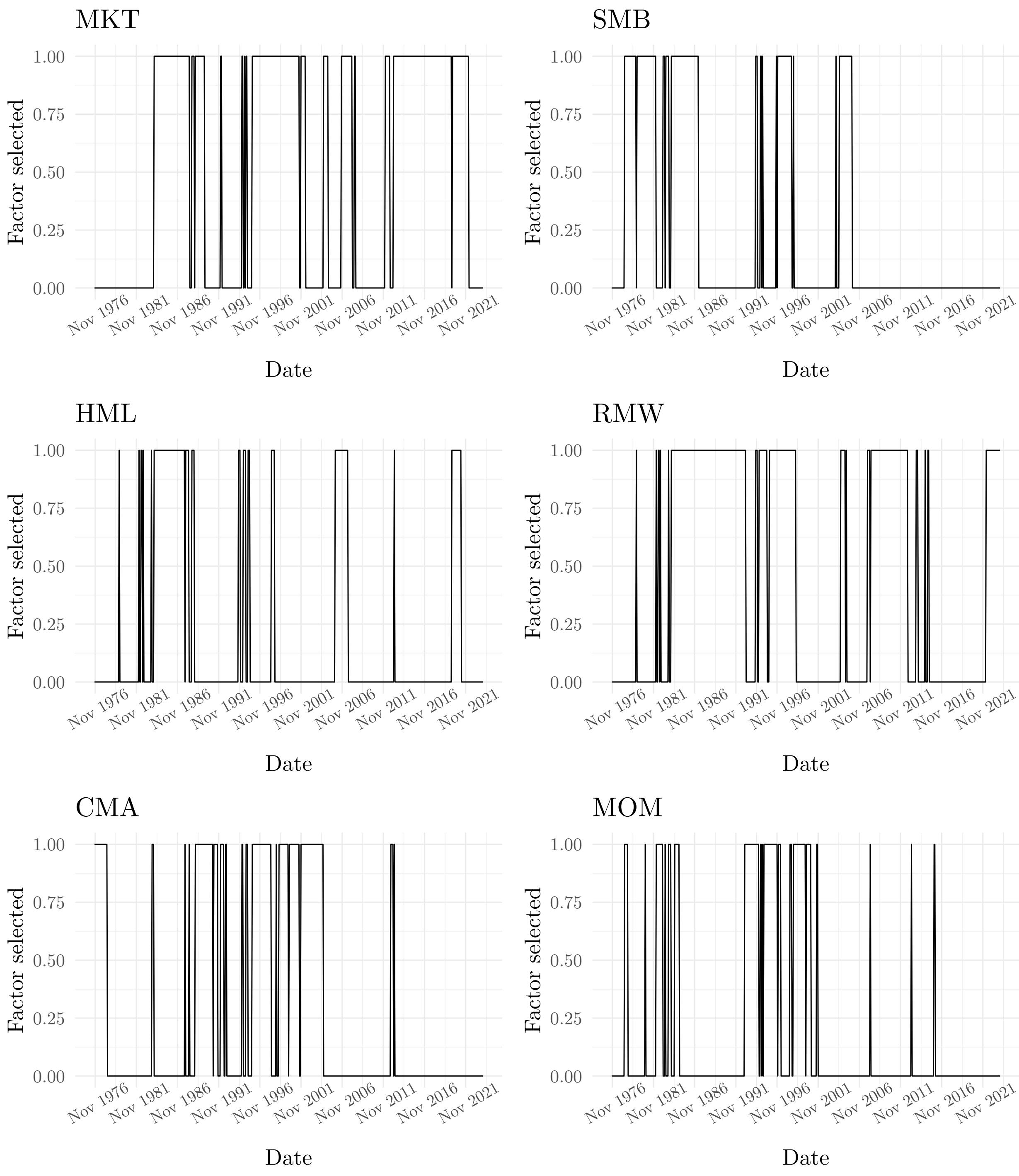


Figure 1: Each subfigure displays whether the factor is selected (indicator variable equals one) or omitted (indicator variable equals zero) at a given point in time, estimated recursively from our methodology that precludes breaks using a five-year fixed rolling window length.

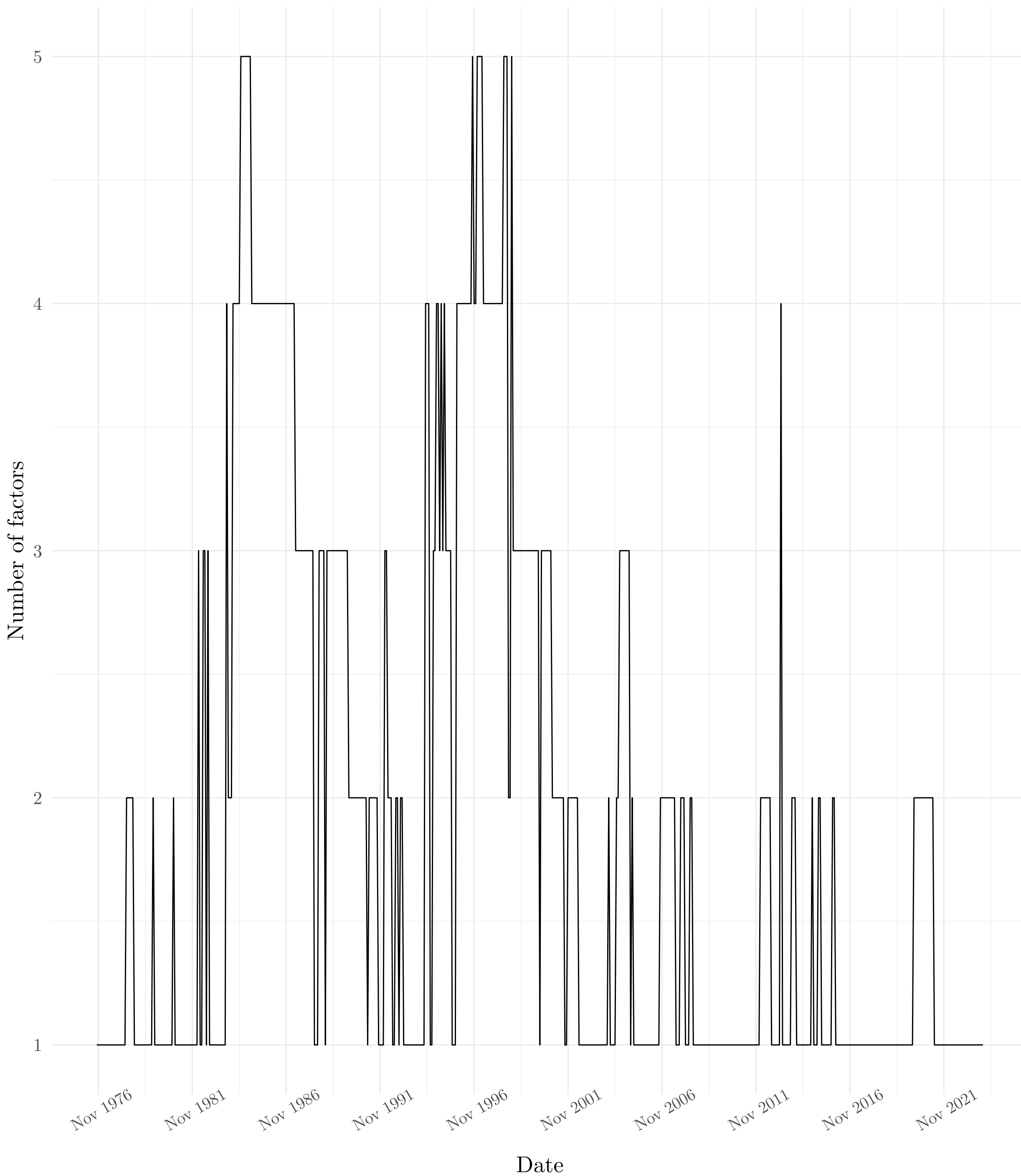


Figure 2: This Figure displays the total number of factors selected at a given point in time, estimated recursively using our methodology that precludes breaks and a five-year fixed rolling window length.

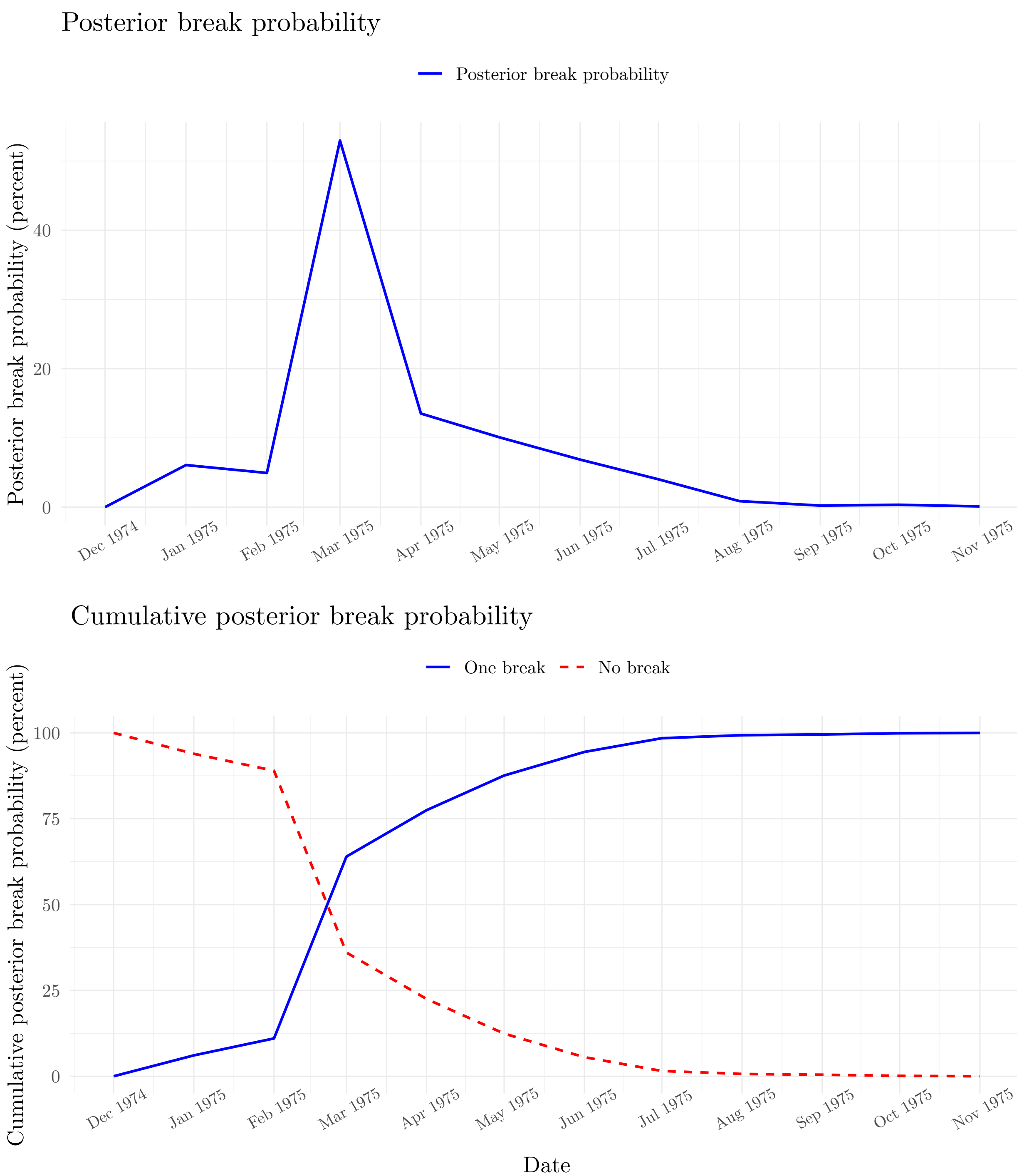


Figure 3: The top panel of this figure displays the posterior probability (measured in percent) that the first break identified by our model occurs in a given month. The lower panel displays the corresponding cumulative posterior probability (measured in percent) that our model has identified no breaks (red dotted line) or one break (blue solid line).

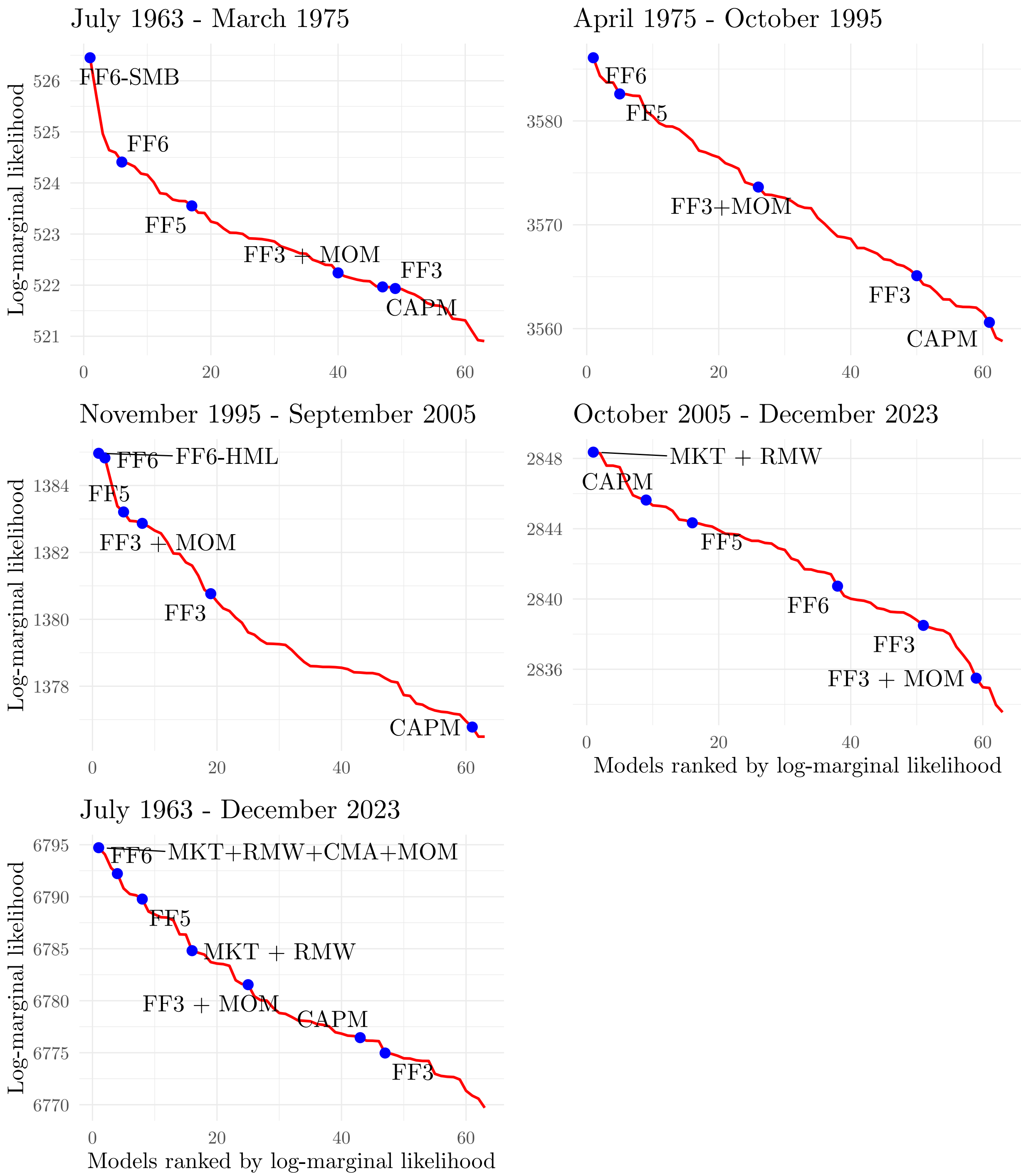


Figure 4: The top four windows of this Figure display, for each of the four regimes, the log marginal likelihoods for each of the 63 possible models which are ranked based on log marginal likelihood from best (left) to worst (right). The best model (top-left circle in each panel) and other models of interest are colored blue. The lower left panel displays the same information for the model that precludes breaks.

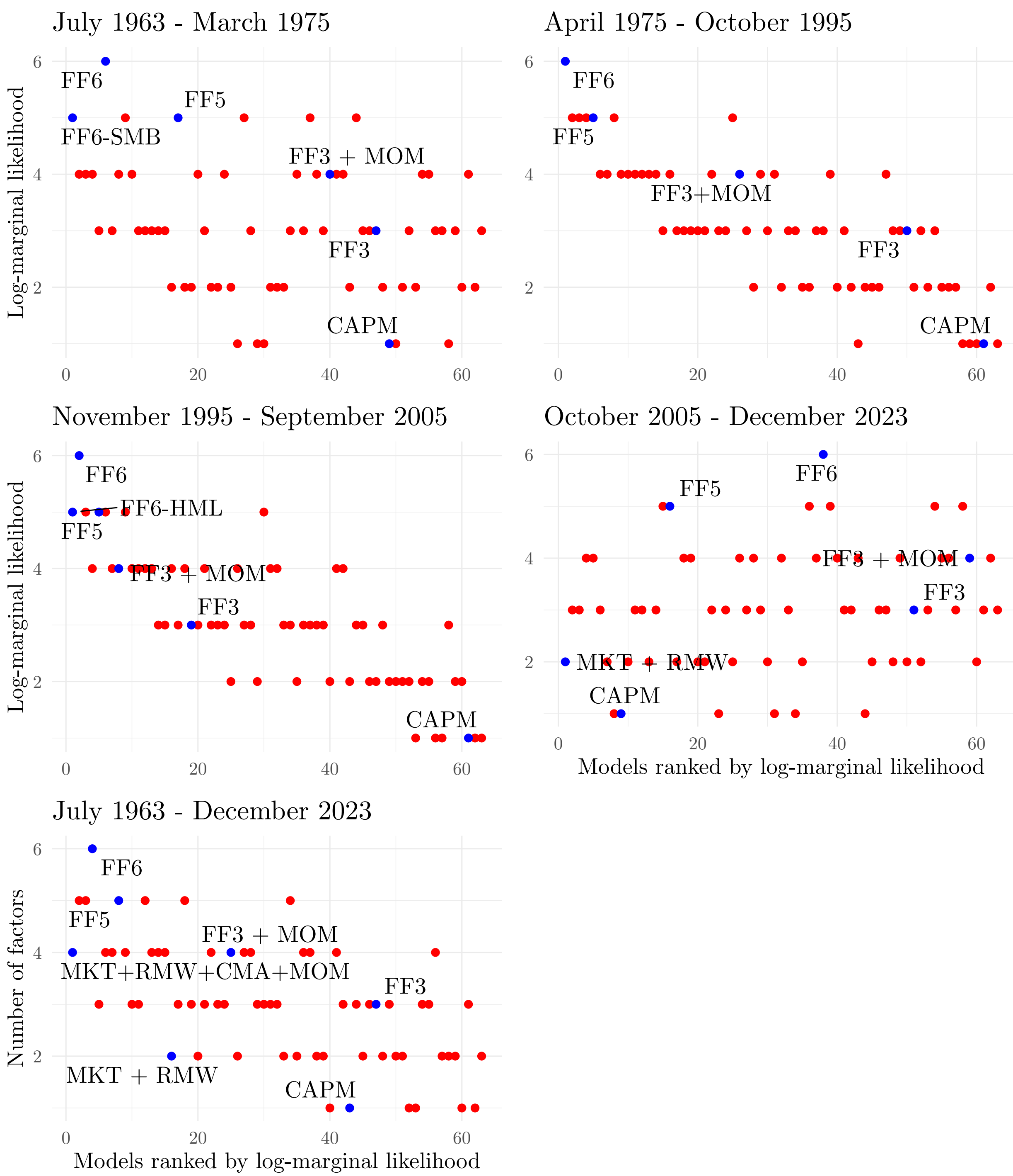


Figure 5: The top four windows of this Figure display, for each of the four regimes, the number of factors in each of the 63 possible models which are ranked based on log marginal likelihood from best (left) to worst (right). The best model and other popular models are displayed in blue. The lower left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

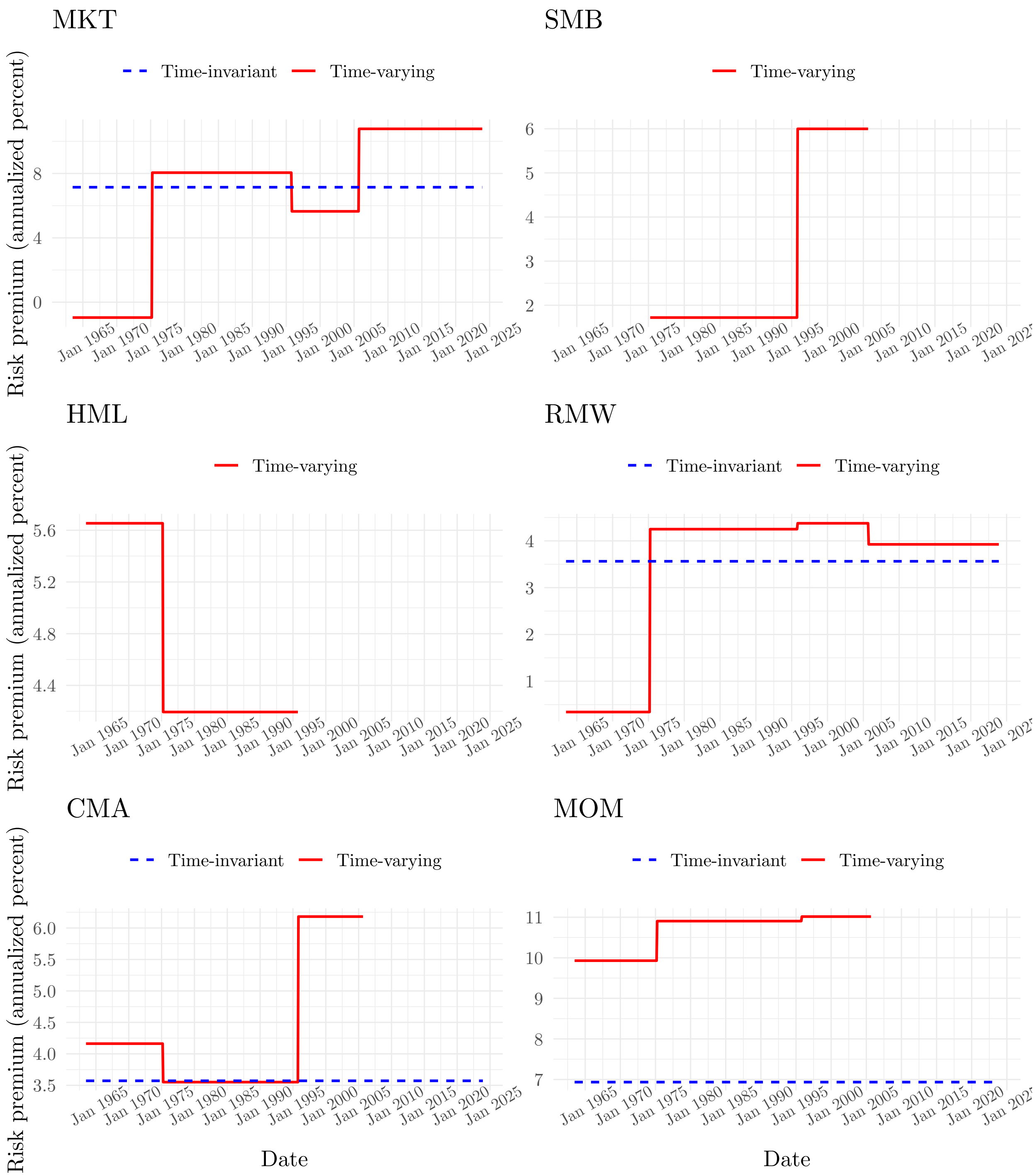
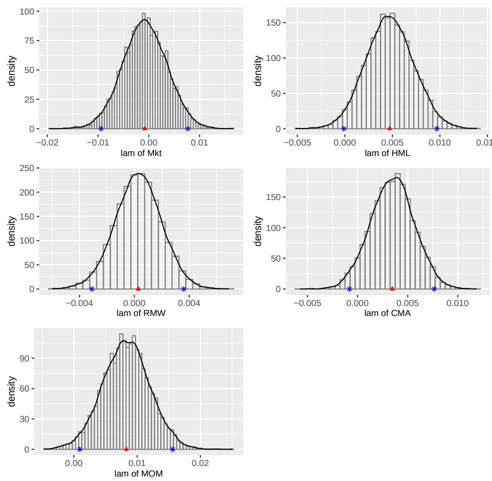
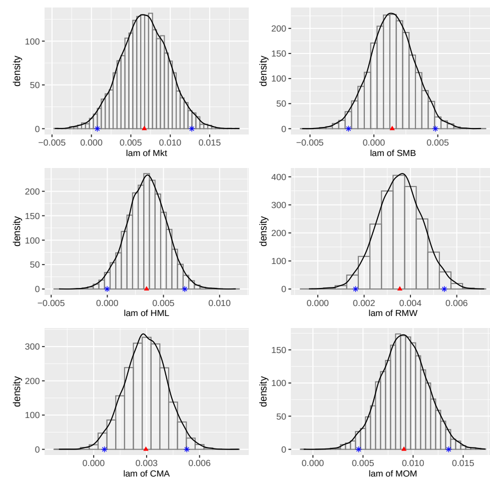


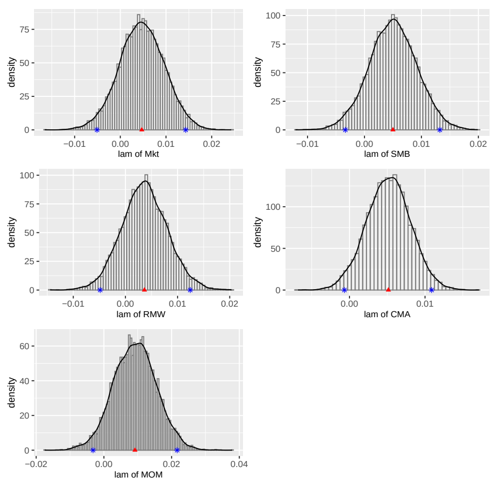
Figure 6: Each window of this Figure displays the time-varying (solid red line) and time-invariant (dashed blue line) risk premia estimates (expressed in annualized percent) for the corresponding factor labelled in the subcaption. The time-varying estimates are from the selected model in each regime and the time-invariant estimates are from the same model that precludes breaks. Parts of the solid red line that are 'missing' correspond to that factor not being selected in the given regime. Similarly, the dashed blue line is omitted for HML and SMB altogether since they are not selected in the model that precludes breaks.



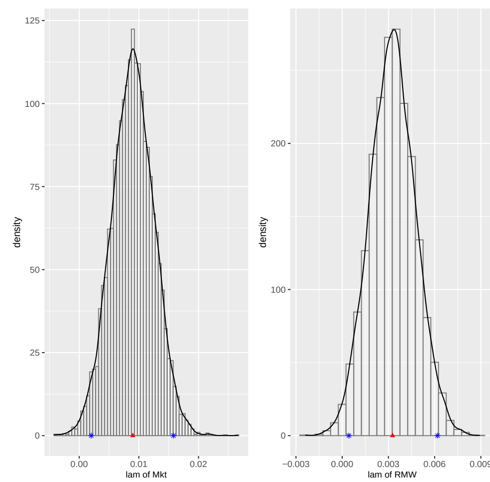
(a) Jul 1963 - Mar 1975



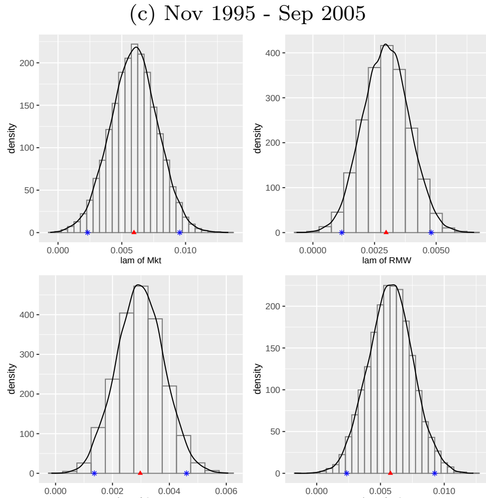
(b) Apr 1975 - Oct 1995



(c) Nov 1995 - Sep 2005



(d) Oct 2005 - Dec 2023



(e) Jul 1963 - Dec 2023

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 lower left panel displays the same information for the model that precludes breaks.

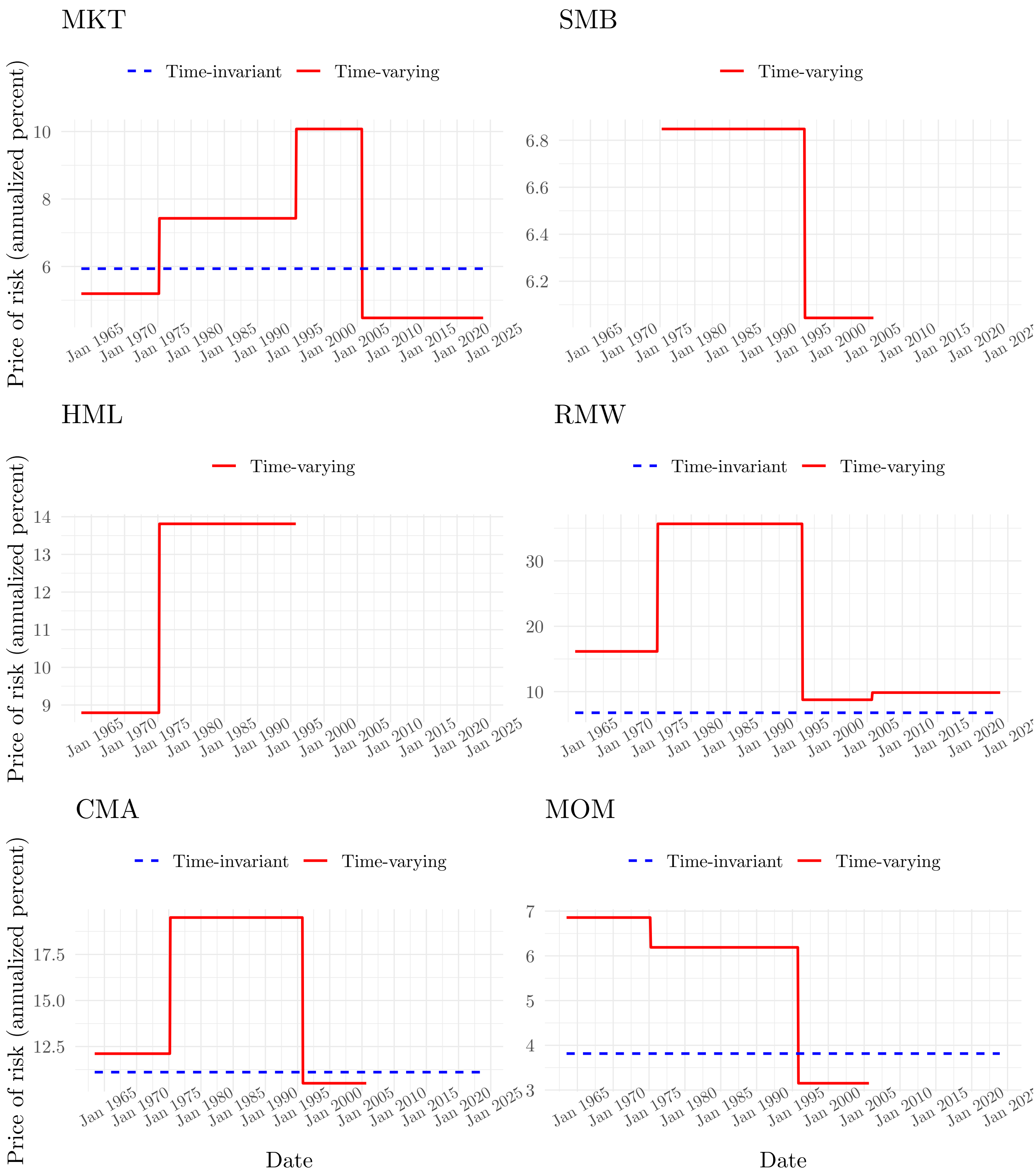


Figure 8: Each window of this Figure displays the time-varying (solid red line) and time-invariant (dashed blue line) price of risk estimates (expressed in annualized percent) for the corresponding factor labelled in the subheading. The time-varying estimates are from the selected model in each regime and the time-invariant estimates are from the same model that precludes breaks. Parts of the solid red line that are 'missing' correspond to that factor not being selected in the given regime. Similarly, the dashed blue line is omitted for HML and SMB altogether since they are not selected in the model that precludes breaks.

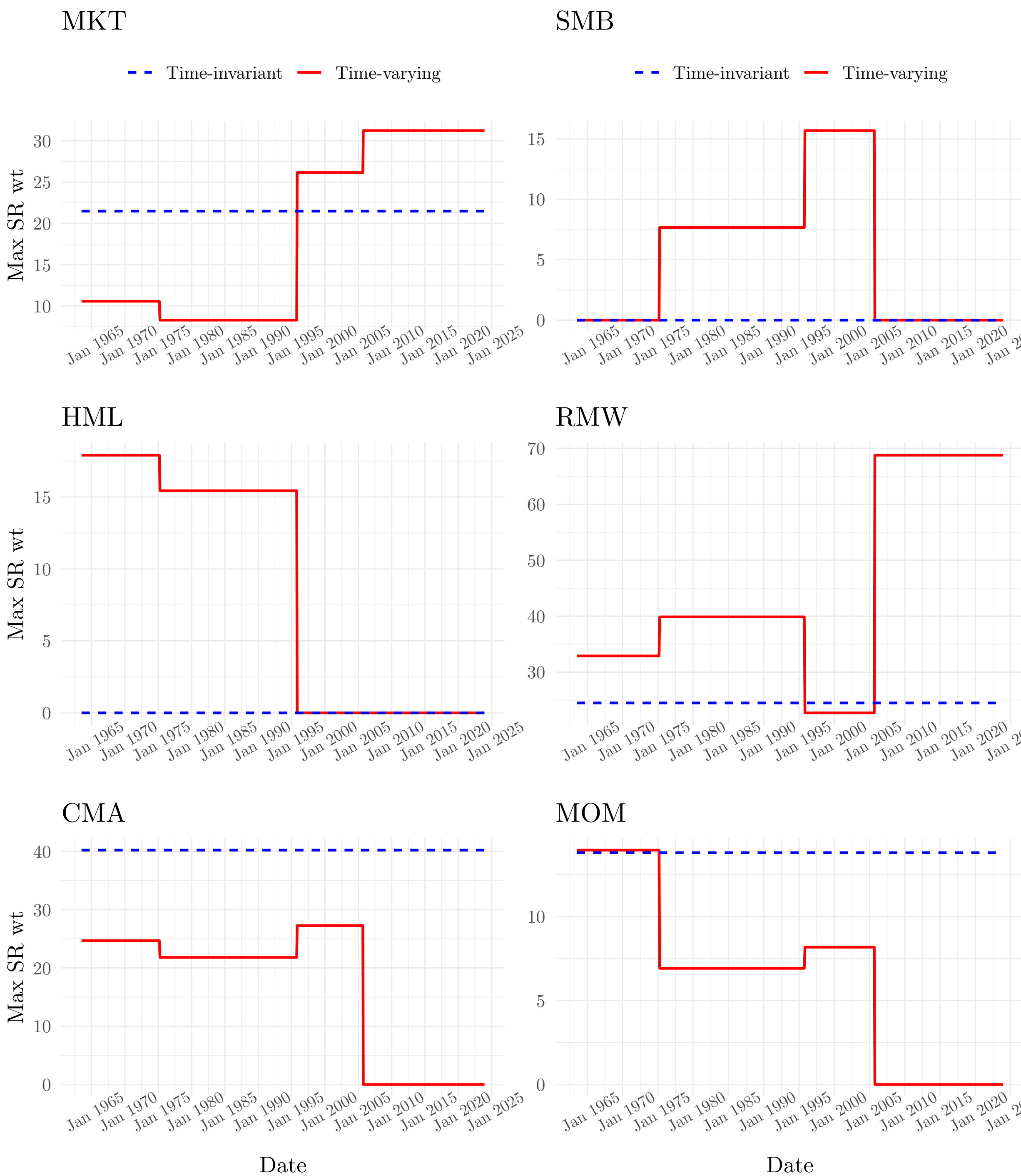


Figure 9: Each window of this Figure displays the time-varying (solid red line) and time-invariant (dashed blue line) estimated weights (expressed in percent) for the corresponding factor labelled in the subtitle in the maximum Sharpe ratio portfolio constructed from the selected factors. The time-varying estimates are from the selected model in each regime and the time-invariant estimates are from the same model that precludes breaks.