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.

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

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).

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：https://d.book118.com/866200202214010142