Uncommon Factors and Asset Heterogeneity *

Lin William Cong Guanhao Feng Jingyu He Junye Li

First version: Sep. 2022; this version: Aug. 2024.

Abstract

To maintain sparsity for economic interpretation and computational feasibility while exploiting the stability and superior empirical performance of large models (e.g., Giannone et al., 2021; Kelly et al., 2023), we introduce a general tree-based AI framework that jointly clusters observations and estimates heterogeneous local models (with Bayesian priors) for analyzing panel data with potential grouped heterogeneity. We derive analytical marginal likelihoods as global split criteria in this Bayesian Clustering Model (BCM), to incorporate economic guidance, address parameter and model uncertainties, and prevent overfitting. We apply BCM to asset pricing and estimate uncommon-factor models for data-driven asset clusters and macroeconomic regimes. We find (i) cross-sectional heterogeneity linked to (non-linear interactions of) return volatility, size, and value, (ii) structural changes in factor relevance over time primarily predicted by market volatility and valuation, and (iii) MKTRF and SMB as common factors and multiple uncommon factors across characteristics-managed-market-timed clusters. BCM helps explain volatility- or size-related anomalies, validate within-group tests, and mitigate the "factor zoo" problem. Overall, BCM outperforms benchmark common-factor models in pricing and investments in U.S. equities, e.g., attaining out-of-sample cross-sectional R^2s exceeding 25% for multiple clusters and Sharpe ratio of tangency portfolios tripling that obtained from 5×5 ME-B/M portfolios.

Key Words: Asset Pricing, Factors, Grouped Heterogeneity, Panel Trees, Sparsity, Structural Breaks.

JEL Classification: C11, C38, G11, G12.

*We thank Rohit Allena (discussant), Doron Avramov, Yi Cao (discussant), I-Hsuan Ethan Chiang (discussant), Siddhartha Chib, Tarun Chordia, John Cochrane, Darrell Duffie, Jianqing Fan, P. Richard Hahn, Cam Harvey, Zhiguo He, Yael Hochberg, Yongmiao Hong, David Hirshleifer, Bob Jarrow, Fuwei Jiang (discussant), Serhiy Kozak, Sophia Zhengzi Li, David Ng, Andrew Patton, Markus Pelger, Xiao Qiao, Alberto Rossi, Olivier Scaillet, Michael Sockin, Oleg Sokolinskiy (discussant), Artem Streltsov, Daniel Titman, Fabio Trojani, Junbo Wang (discussant), Dacheng Xiu, Mao Ye, Guofu Zhou, and seminar and conference participants at BlackRock, CityU HK, Columbia, Cornell, CUHK, 2024 China Fintech Research Conference, EasternFA, 2024 EFMA Annual Meeting, 3rd Frontiers of Factor Investing Conference, Harvest Fund 25th Anniversary Ceremony, HKU, HUST, 4th International FinTech Research Forum (RUC), KAIST Digital Finance Conference, Macquarie University, MFA, Mid-South DATA Conference, NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics, ESIF Economics and AI+ML Meeting (Cornell), NYU Courant, Oxford, Peking University Guanghua, PHBS, Princeton, 2023 Tongji Finance Symposium, Rice, Rutgers Business School, 6th Shanghai Financial Forefront Symposium, Singapore National University, Stanford MS&E, SWFA, Tsinghua University SEM, UCAS, University of Geneva, USC Marshall, UT Austin, 4th Workshop on Big Data Econometric Theory and Application, Xiamen University, and 2023 XJTLU AI and Big Data in Accounting and Finance Conference for invaluable comments and discussions. This paper subsumes results in manuscripts titled "Uncommon Factors for Bayesian Asset Clusters" and "Uncommon Factors and Asset Heterogeneity in the Cross Section and Time Series." We thank Ripple's UBRI for research support as well as Yuanzhi Wang and Qianshu Zhang for outstanding research assistance. Cong (Corresponding, e-mail: will.cong@cornell.edu) is at Cornell University and NBER; Feng (E-mail: gavin.feng@cityu.edu.hk) and He (E-mail: jingyuhe@cityu.edu.hk) are at City University of Hong Kong; Li (E-mail: li_junye@fudan.edu.cn) is at Fudan University.

1 Introduction

Sparse models, widely adopted in economics, utilize a small set of predictors or explanatory variables to achieve estimation tractability and model interpretability (see, e.g., Tibshirani, 1996; Belloni, Chen, Chernozhukov, and Hansen, 2012). This modeling approach assumes a *globally sparse* structure, with a small set of common covariates applied universally to observations. However, potential grouped heterogeneity in the cross section and time series poses an empirical challenge: a model with universal and sparse covariates may not be accurately estimated and often describes only certain episodes or observation groups. For example, relevant predictors can change over time (e.g., due to structural breaks), and parameters estimated from the entire sample may be inaccurate for a subset of observations. Consequently, conventional sparse models may not be stable or effective in empirical predictions (Giannone et al., 2021). This contrasts starkly with the large/overparameterized models that are gaining popularity because of their superior performance and greater generalizability (e.g., Neyshabur et al., 2019; Hastie et al., 2022; Kelly et al., 2023).

To account for unknown heterogeneity groups and develop empirically effective sparse models, we propose an alternative approach to learn from the data any grouped heterogeneity and select the corresponding local sparse models that can be accurately estimated, interpretable, effective on economic big data, and in performance as good as recent large machine learning or AI models that lack economic interpretability. Our Bayesian Clustering Model (BCM) tackles this new econometric problem of jointly selecting variables (and thus the model) on the right-hand side and clustering observations on the left-hand side of regressions. It clusters individual observations (e.g., asset returns) into heterogeneous groups with an economically guided panel tree while estimating heterogeneous local models (e.g., uncommon factors for pricing assets in each cluster) through Bayesian variable selection.

Although BCM generally applies to panel data analysis, we focus on its application to empirical asset pricing, a setting where grouped heterogeneity is prevalent, and sparse modeling that accommodates grouped heterogeneity better answers existing

questions and enables new analyses with theoretical and practical implications. The literature routinely uses common factor models (i.e., Fama and French, 1993, 2015) to understand time-series co-movement and cross-sectional variation of returns. Sparse modeling (e.g., Freyberger et al., 2020; Feng et al., 2020) presents a potential solution to the "factor zoo" problem (Cochrane, 2011). However, most empirical studies evaluate and select factors based on their global performance on ad hoc test assets, overlooking potential grouped heterogeneity that causes failures in accurately pricing certain assets or in certain regimes. For example, models without the momentum factor cannot explain return spreads of recent winners and losers, while only momentum-sorted portfolios have non-trivial exposures to the factor (Carhart, 1997). Macroeconomic environments undergo structural changes as well (Nagel and Singleton, 2011), with investors learning over time and in the aggregate, which affects the return dynamics.

For all these reasons, the "correct" factor model may well manifest itself differently across (pre-specified or unknown) asset groups and macroeconomic regimes, rendering the search for it challenging.[1] To tackle the challenge, BCM exploits financial panel data to determine "which risk factors explain what asset returns and when." It learns an interpretable and parsimonious description of potential (i) cross-sectional heterogeneity based on high-dimensional asset characteristics and (ii) structural breaks and asset-pricing regimes instrumented by macroeconomic predictors.

Beyond common factors, BCM uses a sparse set of uncommon factors to capture cluster-specific asset return co-movements. Ultimately, BCM produces a small number of macro-instrumented regimes and characteristics-managed asset clusters, and determines a few factors in each cluster that best explain or predict its return dispersion and dynamics. BCM also clarifies that grouped heterogeneity (in factor-model fitting rather than ad hoc characteristics correlated with returns) mitigates the errors-in-variables problem (e.g., Berk, 2000) and validates the empirical procedure involving sorting for within-group tests.[2]

[1] Due to capital controls, investment mandates, limited participation, market segmentation, or incompleteness, investors vary across regimes, industries, countries, and asset classes (Levy, 1978; Merton, 1987). Investors may find certain factors relevant only for certain assets during specific periods. These issues have implications for trading, asset prices, and risk premia (e.g., Black, 1974; Errunza and Losq, 1985), especially with limited arbitrage or slow-moving capital.

[2] Even though the total variance is distributed across heterogeneity groups (and thus lower in any

More generally, BCM introduces several conceptual and methodological innovations. First, our approach, along with Cong et al. (2022), is the earliest to customize decision trees for analyzing financial panel data and goal-oriented clustering (neither supervised nor unsupervised). The guidance from improving the fitness of an economic model complements pure statistical correlations in the data and guards against the well-known overfitting problem of tree-based models by using a natural global split criterion in the joint marginal likelihood. BCM can have sparsity in both the selected variables and the determinants for the clusters (e.g., a small set of characteristics or macroeconomic variables in the asset pricing application), in contrast to rotation- or shrinkage-based methods (e.g., PCA or Ridge for latent factors) which are not sparse in primitive features. Second, the spike-and-slab prior (also adopted in Giannone et al., 2021) allows us to enjoy the usual benefits of a Bayesian approach, such as accounting for parameter uncertainty, prior economic knowledge, and model uncertainty. This approach maintains the advantages of dense modeling, which is also emphasized by Avramov et al. (2023), mitigating model specification errors. Third, in the context of asset pricing, we offer a new (non-linear) way to combine information from characteristics and the large number of factors documented in the literature, generalizing the discussion beyond low-dimensional linear settings (e.g., Daniel and Titman, 1997).

Empirical Highlights. In our BCM application to empirical asset pricing, we examine the entire universe of U.S. individual stocks in the sample period from 1972 to 2021. The cross section and time series of returns, $r_{i,t}$, are the panel data observations, and we consider a candidate set of widely used risk factors f_t as potential variables for selection. Our baseline specification investigates cross-sectional heterogeneity based on 20 representative asset characteristics $z_{i,t-1}$, which are the features for clustering.

Return variance, market equity, and book-to-market are the most important features describing the grouped heterogeneity in characteristics-managed clusters. MKTRF (Market factor) and SMB (Size factor) are common factors with high selection

particular group), the distribution of measurement error in beta within a group generally differs from that in the whole sample due to heterogeneity allowed in (1). Furthermore, the total variance is also used to estimate multiple factors, which means that the relative noise within a heterogeneity group for estimating selected factors can be lower than in the whole sample considering all factors. BCM thus enables one to use information from within-group tests.

probabilities.[3] RMW and ROE (i.e., the two Profitability factors) are selected in only three asset clusters with a probability above 0.7, while CMA and IA (i.e., the two Investment factors) have low selection probabilities in most clusters. Notably, BCM creates a unique asset cluster (high stock return variance, low earnings-to-price, low momentum, and low market equity) that comprises approximately 4% of all stocks, with almost no exposure to any factors.

Although most factors (18 out of 20) show high selection probabilities when evaluated against the aggregate stock universe, some become "uncommon" and relevant only for certain clusters once group heterogeneity is considered. Some even become useless as assets are partitioned into finer clusters. Notably, we find that several well-known factors (such as HML for value, RMW for profitability, and UMD for momentum) are only partially useful during the expansion of the cross section of leaf clusters. Furthermore, BCM improves asset pricing performance with clustering and cluster-specific factor models. Benchmark factor models, such as the Fama-French Five Factor Model (FF5), show larger pricing errors than cluster-specific factor models in all asset clusters resulting from BCM. The cluster-specified factor model provides better investment guidance. For instance, an agnostic Bayesian investor to risk factors can earn an out-of-sample monthly CAPM alpha of 0.86%.

We further expand our investigation to consider the heterogeneity of factor structures in asset returns over time. Unlike conventional regime-switching models that detect structural breaks based on calendar time, the BCM identifies regimes that are "instrumented" by macroeconomic predictors and guided by the economic relevance of asset pricing. BCM identifies the three most prominent macroeconomic regimes that alternate over time: (i) high volatility, (ii) high valuation and non-high volatility, and (iii) low valuation and non-high volatility. Under each regime, firm characteristics such as return variance and size are critical to further capturing cross-sectional heterogeneity. Remarkably, the selected firm characteristics and splitting thresholds are the same at the top two layers under the low- and high-valuation regimes. However, the resulting asset clusters have distinct sets of selected factors. The sparsity model

[3]The list of factors and their acronyms are presented in Section 3.

for grouped heterogeneity in time series and cross sections significantly improves the performance in pricing individual asset returns, indicating that oversimplified global sparsity assumptions might lead to information loss.

Literature. Our paper contributes to the extensive literature on sparse (factor) modeling and emerging studies using large models. Lasso and its other penalized methods (e.g., Tibshirani, 1996; Fan and Li, 2001; Belloni et al., 2012) are widely used for sparse variable selection, including in return forecasts (e.g. Chinco et al., 2019). Belloni et al. (2014) develop post-selection inference methods that control high-dimensional variables, which Feng et al. (2020) adapt to asset pricing. Bayesian variable selection features either the spike-and-slab prior (Mitchell and Beauchamp, 1988; George and McCulloch, 1993) or the horseshoe prior (Carvalho et al., 2010), and has recently been applied to cross-sectional asset pricing (e.g., Bryzgalova et al., 2023).

Dense (or large) models, in contrast, assume many covariates to be useful, even if some only have small effects, and include principal component analysis (Stock and Watson, 2002), Ridge regularization (Kozak et al., 2020), and large (deep learning) models (e.g., Cong et al., 2020). Fan et al. (2022) and Kelly et al. (2023) explain why they perform extremely well empirically in return predictions. Our key innovation lies in effectively combining elements from both fields to achieve interpretability, estimation tractability, and the capacity to handle big data and mitigate model misspecification simultaneously.

Our paper, along with Cong et al. (2022), is among the first to develop goal-oriented clustering, which searches in a flexible clustering space in a data-driven and yet economically guided manner. Previous studies have analyzed heterogeneous asset groups using only heuristics, statistical correlations, or the researchers' expertise. For example, different factors are adopted for assets in different countries, asset classes, business cycles, or groups with strong comovements (see, e.g., Foerster and Karolyi, 1999; Griffin, 2002; Bekaert et al., 2009; Hou et al., 2011; Chaieb et al., 2021).[4] Heterogeneity clusters also appear in causal inferences and controlled randomized trials

[4]While not clustering assets or studying uncommon-factor models, Jarrow and Protter (2016) were among the earliest to point out the possible presence of a vast number of risk factors —- a justification for dense modeling—and that different assets rely on different risk factors.

(e.g., Bugni et al., 2022), but again the clusters are typically exogenously specified by researchers.[5]

Closely related to ours by also analyzing endogenous grouped heterogeneity in financial markets, Ahn, Conrad, and Dittmar (2009) use unsupervised clustering based on return correlations, and Patton and Weller (2022) build on Bonhomme and Manresa (2015) to generalize K-means to group assets based on within-group slopes and averages, and find risk-price heterogeneity pervasive and important. Our paper distinguishes itself by emphasizing interpretability, and jointly selecting variables while clustering observations in the panel data. Our modeling of grouped heterogeneity in asset prices is also partially motivated by the weak factor problem (e.g., Giglio et al., 2022), which BCM circumvents by clustering assets for heterogeneous factor models.[6]

Our paper further contributes to conditional asset pricing models that account for time variation and regime switches. Empirical evidence suggests that the underlying asset pricing models change over time (e.g., Pastor and Stambaugh, 2001; Ang and Bekaert, 2002; Smith and Timmermann, 2021, 2022). While models of time-varying coefficients have been applied in conditional asset pricing (e.g., Nagel and Singleton, 2011; Gagliardini et al., 2016; Cui et al., 2023), time-varying model specifications with regime switching remain underdeveloped. Guided by asset pricing model fitness, BCM handles time-series breaks, identifies regimes using macroeconomic predictors, and selects time-varying specified factor models for different regimes (clusters).

In terms of methodology, BCM integrates decision trees and Bayesian statistical methods into a unified framework for sparse modeling with grouped heterogeneity. Bayesian methods have long been applied in empirical finance (see, e.g., Avramov and Chordia, 2006; Barillas and Shanken, 2018; Chib and Zeng, 2020; Avramov et al., 2023; Chib et al., 2023), and are argued to be the ultimate solution to asset pricing (see the AFA presidential address by Harvey, 2017). We use spike-and-slab priors to estimate

[5] In statistics, Ibragimov and Müller (2010) develop a general strategy for conducting inference in potentially heterogeneous and correlated data. While the paper justifies the popular Fama-MacBeth method, the focus is on heterogeneity in variance correlations. More recently, Canay et al. (2021) derive bootstrapping analyses and asymptotic properties for a fixed set with a small number of heterogeneous clusters.

[6] Various methods address weak identification/model misspecification issues (e.g., Kleibergen, 2009; Gospodinov et al., 2014). Similarly to ours, Giglio et al. (2022) suggest that factor strength depends on selecting test assets and propose using supervised PCA to choose them.

the marginal likelihood of asset returns instead of finding common linear factors or estimating risk premia for fixed test assets. Importantly, BCM computes efficiently and offers insights and performance improvement in pricing and investments by combining information on characteristics and factors, rather than enumerating all variable combinations.

Tree-based machine learning is gaining attention in asset pricing due to its ability to graphically represent nonlinear interactions and effectively model high-dimensional characteristics (e.g., Rossi and Timmermann, 2015; Gu et al., 2020; Bianchi et al., 2021; Van Binsbergen et al., 2023). Recently, Bryzgalova et al. (2022) apply shrinkage to prune a forest of shallow trees to select effective basis portfolios. Cong et al. (2022) introduce "panel trees" to generalize security sorting for deriving basis portfolios and common latent factors. Like Cong et al. (2022) but unlike Bryzgalova et al. (2022), BCM exploits top-down tree growth for economically guided, goal-oriented search, thus being able to explore a much larger universe of potential clusters efficiently, allowing deep asymmetric interactions of asset characteristics and macroeconomic variables. Relative to ensemble models (XBART, He et al., 2019; He and Hahn, 2021), BCM focuses on clustering using natural global split criteria that prevent overfitting and preserve single-tree interpretability (e.g., Athey and Imbens, 2019).

The remainder of the article is organized as follows. Section 2 presents the BCM framework. Section 3 describes the data used for our empirical analysis. Section 4 presents the baseline empirical findings, including asset pricing and investment performance. Section 5 extends BCM to account for regime switches and structural changes relevant to asset pricing. Section 6 concludes.

2 BCM for Panel Data Analysis

BCM, as a general modeling framework, combines tree-based clustering and Bayesian variable selection. When applied to economic panel data, it generates heterogeneity groups in the cross section and/or time series. It then selects potentially uncommon variables for modeling each group while keeping the overall architecture sparse and interpretable. Section 2.1 introduces a model of panel data generation (e.g., a model

of asset returns), Section 2.2 discusses the variable/model selection (e.g., factors) in each cluster using Bayesian Spike-and-Slab priors, and then Section 2.3 presents the algorithm for clustering observations and fitting a heterogeneous model (e.g., uncommon factor model) for each leaf cluster. Given our application focus, for the remainder of the paper, we mostly use terminologies in asset pricing, though they have general counterparts in panel data analysis in other fields as well.

2.1 Panel Data with Grouped Heterogeneity

We consider the underlying data-generating process of a potentially unbalanced and large panel of observations. We denote the excess return (dependent outcome) of the individual asset (unit) i at time t by $r_{i,t}$, where the integer i ranges from 1 to n_t, the number of assets (units) available at time t. We let $\mathbf{z}_{i,t-1}$ denote the vector of M characteristics of the i-th asset (unit) in the time period $t-1$, $\mathbf{x}_{t-1}$ a vector of macro variables (common controls), and $\mathbf{f}_t$ a vector of K candidate factors (explanatory variables).

We first introduce the complete model of the data-generating process with clusters *a posteriori*, and provide further details on clustering in Section 2.3. Suppose asset returns exhibit grouped heterogeneity in the cross section and/or time series, in that they follow a potentially different factor structure (as reflected in the coefficients on the factors) across the groups (i.e., clusters):

$$
\begin{aligned}
r_{i,t} &= A(\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1}) + B(\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1})^{\mathsf{T}} \mathbf{f}_t + \epsilon_{i,t}, \\
\textit{where} \qquad A(\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1}) &= \sum_{j=1}^{J} \mathbb{1}_{\{T(\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1})=j\}} \alpha_j, \\
B(\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1}) &= \sum_{j=1}^{J} \mathbb{1}_{\{T(\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1})=j\}} \boldsymbol{\beta}_j(\mathbf{z}_{i,t-1}), \qquad (1) \\
\boldsymbol{\beta}_j(\mathbf{z}_{i,t-1}) &= \mathbf{b}_{j,0} + \mathbf{b}_{j,1}^{\mathsf{T}} \left(I_K \otimes \mathbf{z}_{i,t-1} \right), \\
\epsilon_{i,t} \overset{\text{independent}}{\sim} N(0, \sigma_{i,t}^2), \quad \sigma_{i,t}^2 &= \sum_{j=1}^{J} \mathbb{1}_{\{T(\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1})=j\}} \sigma_j^2,
\end{aligned}
$$

where $\mathbf{f}_t$ is a vector of K pre-specified factors (the candidate pool from which we perform Bayesian factor selection), I_K, a $K \times K$ identity matrix, $\mathbb{1}(\cdot)$, the indicator func-

tion, and $\otimes$, the Kronecker product. The clustering function $T(\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1})$ describes the latent grouped heterogeneity, which is assumed to be a deterministic function of asset characteristics $\mathbf{z}_{i,t-1}$ and/or macro variables $\mathbf{x}_{t-1}$. It maps the characteristics and macro variables for each unit observation to one (and only one) group number in $1, 2, \ldots, J$, where J denotes the number of clusters to be estimated.[7] For simplicity, we specify $A(\cdot)$ as a cluster-specific scalar, and $B(\cdot)$ as a $K \times 1$ vector for heterogeneous alpha and beta, respectively. They both depend (in a non-linear way with interactions, see, e.g., Avramov and Chordia, 2006) on the characteristics through clustering rules (i.e., $T(\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1})$).

Generalizing traditional security sorting to allow for interactions among multiple characteristics or even time-series variables, we empirically learn the clustering function $T(\cdot)$ using a customized tree that we introduce shortly. We consider two tree-based specifications of $T(\cdot)$, which then play a crucial role in partitioning the asset space based on firm characteristics and macro predictors through a sequence of split rules (cutpoints). In the baseline, the BCM tree is driven solely by firm characteristics $T(\mathbf{z}_{i,t-1})$ to identify the cross-sectional heterogeneity of asset returns; later, we allow the tree to be driven by both firm characteristics and macro predictors $T(\mathbf{z}_{i,t-1}, \mathbf{x}_{t-1})$ to identify heterogeneity in both the cross section and time series (changes in asset-pricing-related regimes).

2.2 Bayesian Model Selection in a BCM Leaf

We start with the Bayesian model selection in a given cluster (a single leaf node of the tree), i.e., we focus on asset return observations, $r_{i,t}$, in the same j-th cluster for some j. Without worrying about cluster assignment, substituting the dynamics of $A(\cdot)$ and $B(\cdot)$ into the return dynamics, while using the identity $(I_K \otimes \mathbf{z}_{i,t-1})\mathbf{f}_t = \mathbf{f}_t \otimes \mathbf{z}_{i,t-1}$, one gets:

$$r_{i,t} = \alpha_j + \mathbf{b}_{j,0}^{\mathsf{T}} \mathbf{f}_t + \mathbf{b}_{j,1}^{\mathsf{T}} (\mathbf{f}_t \otimes \mathbf{z}_{i,t-1}) + \epsilon_{i,t}. \tag{2}$$

The coefficients vector $\mathbf{b}_{j,0} = \{b_{j,0,k} \mid 1 \leq k \leq K\}$ has length K while $\mathbf{b}_{j,1} = \{b_{j,1,k,m} \mid 1 \leq k \leq K, 1 \leq m \leq M\}$ has length KM, and each element $b_{j,1,k,m}$ denotes the re-

[7] BCM is data-driven in that J and the clustering function are not pre-specified when we take the model to data, but are learned endogenously from the data.

gression coefficient of the interaction term between k-th factor of $\mathbf{f}_t$ and m-th firm's characteristics, $\mathbf{z}_{i,t-1}$.

Group Spike-and-Slab prior. Notice that (2) consists of factor $\mathbf{f}_t$ and interaction terms $\mathbf{f}_t \otimes \mathbf{z}_{i,t-1}$ that are results of the dynamic beta. The number of factors and firm characteristics can be large, resulting in a large number of variables in the model. We perform Bayesian variable selection on factors, assuming that each factor and its interactions with characteristics entering the model jointly, i.e., all interaction terms will be selected once the corresponding factor is selected, and vice versa. As such, we modify the Bayesian spike-and-slab priors (George and McCulloch, 1993) into a *group* spike-and-slab prior, which is a Bayesian counterpart of the group Lasso for joint variable selection.

We assume independent priors for regression coefficients vector $\mathbf{b}_{j,0}$, on which the prior for $\mathbf{b}_{j,1}$ depends. A vector of the latent variables taking the value of 1 or 0 to indicate whether the prior on corresponding coefficient $[\mathbf{b}_{j,0}, \mathbf{b}_{j,1}]$ is in the region of "slab" or "spike," i.e., whether the corresponding variable in (2) are selected in the model or not,

$$\boldsymbol{\gamma}_j = (\gamma_{j,1}, \gamma_{j,2}, \cdots, \gamma_{j,K}, \cdots, \gamma_{j,K+KM}) = (\underbrace{\boldsymbol{\gamma}_{j,\mathbf{f}}^{\mathsf{T}}}_{K\times 1}, \underbrace{\boldsymbol{\gamma}_{j,\mathbf{f}\bullet\mathbf{z}}^{\mathsf{T}}}_{KM\times 1})^{\mathsf{T}}, \tag{3}$$

where to simplify the notation, we decompose the length $K + K \times M$ vector $\boldsymbol{\gamma}_j$ into two components $\boldsymbol{\gamma}_{j,\mathbf{f}}$ and $\boldsymbol{\gamma}_{j,\mathbf{f}\bullet\mathbf{z}}$ corresponding to factors $\mathbf{f}_t$ and all interaction terms $\mathbf{f}_t \otimes \mathbf{z}_{i,t-1}$, and $\gamma_{j,\mathbf{f},k}$ is the k-th element of $\boldsymbol{\gamma}_{j,\mathbf{f}}$. To perform group selection, we assume that the coefficient $b_{j,1,k,m}$ does not have its free indicator but follows the corresponding $\gamma_{j,\mathbf{f},k}$. Therefore $\boldsymbol{\gamma}_{j,\mathbf{f}\bullet\mathbf{z}}$ is uniquely defined once $\boldsymbol{\gamma}_{j,\mathbf{f}}$ is given. This way, the main effect and all interactions of factors are jointly selected or removed from the model.

The prior distributions of the regression coefficients are:

$$
\begin{aligned}
&\pi(b_{j,0,k} \mid \sigma_j^2, \boldsymbol{\gamma}_j) = (1-\gamma_{j,\mathbf{f},k})N(0, \xi_0^2\sigma_j^2) + \gamma_{j,\mathbf{f},k}N(0, \xi_1^2\sigma_j^2),\\
&\pi(b_{j,1,k,i} \mid \sigma_j^2, \boldsymbol{\gamma}_j) = (1-\gamma_{j,\mathbf{f},k})N(0, \xi_0^2\sigma_j^2) + \gamma_{j,\mathbf{f},k}N(0, \xi_1^2\sigma_j^2),\\
&\pi(\alpha_{j,0} \mid \sigma_j^2) = N(0, \xi^2\sigma_j^2),\\
&\pi(\sigma_j^2) = \text{inverse-Gamma}(S_0, v_0),\\
&\pi(\boldsymbol{\gamma}_j) = \pi(\boldsymbol{\gamma}_{j,\mathbf{f}}) = \prod_{k=1}^{K} w_k^{\gamma_{j,\mathbf{f},k}}(1-w_k)^{(1-\gamma_{j,\mathbf{f},k})},
\end{aligned}
\tag{4}
$$

for $k = 1, \cdots, K$, and $i = 1, \cdots, M$. We use the continuous version of the spike-and-slab prior for factor selection (George and McCulloch, 1993), a mixture of two normal distributions with different variances. If $\gamma_{j,\mathbf{f},k} = 1$, the corresponding k-th factor $f_{t,k}$ is selected, and the prior is in the "slab" with large ξ_1. Thus, the coefficient is less shrunk. If $\gamma_{j,\mathbf{f},k} = 0$, the k-th factor is not selected, and the prior is in the "spike" with very small ξ_0 to shrink the coefficients towards zero. In addition, ξ is large to reflect almost no shrinkage of alpha.

The tuning parameter w_k represents the prior probability of selecting the corresponding factor ($\gamma_{j,\mathbf{f},k} = 1$). Similar to Avramov (2002), our model incorporates investors' prior beliefs about factor usefulness through the prior parameter values for w_k. For example, an *agnostic* investor with equal uncertainty about the usefulness of factors may choose $w_k = 0.5$, implying an equal prior probability to select or remove a factor. On the other hand, a *skeptical* investor who believes that most factors are useless a priori may select a very small value of w_k (e.g., $w_k = 0.1$) to reflect their low prior belief. Furthermore, we assume the residual variance follows a standard conjugate inverse-Gamma prior distribution.

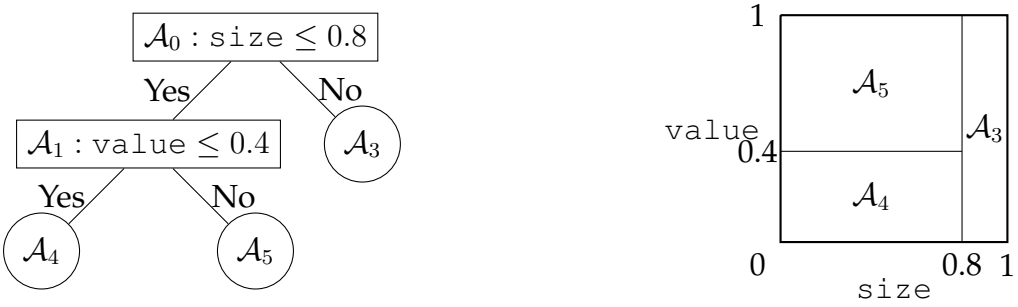
2.3 Clustering by Growing a BCM Tree

With the knowledge of how we select factors (and models) in each cluster, we discuss how the clusters are jointly determined. We first introduce a global split criterion for growing the tree based on the marginal likelihood of the model in (2), then the Gibbs sampler for inference of parameters in the model. A Tree is essentially

a sequence of split rules (cutpoints) that partitions the space of characteristics into many hyper-rectangular regions (leaf nodes). By splitting using firm characteristics, the BCM tree can learn cross-sectional heterogeneity as assets with different characteristic values are partitioned into different sides. Similarly, splitting using macro-variables allows the tree to learn time-series heterogeneity, such as regime switching, as different periods with varying macro conditions are partitioned into different sides. For simplicity, the following section expresses a split rule $\{z \leq c\}$ as $\tilde{c}$, but it should be noted that the variable z can be either firm characteristics or macro predictors. The split rule $\{z \leq c\}$ at each intermediate node of the tree checks whether a specific variable (z) is smaller than a threshold (c). Figure 1 illustrates.

Figure 1: **Decision Tree Partitioning Two Dimensional Characteristics Space.**

Left: A decision tree with two splits and three leaves. Right: Corresponding partition plot for the predictor space spanned by `size` and `value`. This tree partitions the entire space into three rectangular $\mathcal{A}_3$, $\mathcal{A}_4$, $\mathcal{A}_5$.



Suppose a new asset return with characteristics (e.g., size and value) comes; we start by passing it from the (top) root node $\mathcal{A}_0$, check all the split rules, and then assign it to one and only one of the three (bottom) *leaf nodes*, or *clusters*. All leaf nodes constitute a partition of the entire space, as shown in the right panel. We call nodes other than the root and final leaves intermediate nodes (with split rules), such as $\mathcal{A}_1$ in Figure 1. J is then the number of final leaves.

A BCM tree grows iteratively. In the baseline specification, before the first split, the entire cross section of the asset returns is at the root of the tree, $\mathcal{A}_0$. BCM starts from the root node, searches for the optimal split rule that maximizes the objective (the split criterion), among all possible candidates defined by various characteristics or macro variables and thresholds, and creates two leaf nodes (bottom nodes). Then

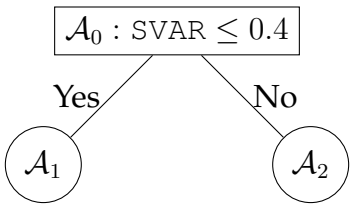
we evaluate the leaf nodes to find the second optimal split rule. Besides the split rule candidates, we add one more option of "stop splitting" at each iteration, with a penalty of the tree size to help the tree stop once there is no desirable split rule. It prevents growing the tree too large (thus less interpretable) and alleviates the over-fitting problem. The iteration terminates once all leaf nodes reach "stop splitting" or other stopping conditions, such as reaching the pre-specified maximum depth of the tree or the minimal number of data observations in a leaf. We next elaborate on the procedures and include the pseudo codes in Algorithm 1 in Appendix I.

2.3.1 First Split

We explore all split rule candidates, including every pair of firm characteristic (and macro variable) and split threshold, to partition the cross section (panel). All characteristics are normalized cross-sectionally between $[0, 1]$ based on the percentiles. Figure 2 presents one split rule candidate, which partitions the sample into the left and right child leaf nodes, $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively, based on whether individual asset's SVAR falls in the bottom 40 percentiles.

Figure 2: **Illustration of one candidate for the first split**

For calculating the split criterion, to search for the best characteristic to split the optimal cutpoint, let us consider one split candidate, $\texttt{SVAR} \leq 0.4$.



The partitioning generally aims to group similar observations into the same leaf to fit a locally sparse model well. To evaluate split rule candidates, the "fitness" of the resulting factor model at each leaf is a natural split criterion. However, the likelihood function of the model in (2) involves unknown parameters that cannot be accurately estimated given the noisy data, which may favor a bad split rule. Instead, we use the closed-form expression of the *marginal likelihood*, where all fitted parameters are integrated out, to address any concerns about parameter uncertainty during tree growth.

Marginal likelihood. Stack all data in the *same cluster* in matrix form, $\mathbf{R}, \mathbf{Z}, \mathbf{F}$, and let $\mathbf{F} \bullet \mathbf{Z}$ denote the stacked matrix of $\mathbf{f}_t \otimes \mathbf{z}_{i,t-1}$. The marginal likelihood of the model at node $\mathcal{A}_0$ is given by:

$$\begin{aligned} p(\mathcal{A}_0) := p(\mathbf{R} \mid \mathbf{Z}, \mathbf{F}) = \int p(\mathbf{R} \mid \mathbf{Z}, \mathbf{F}, \boldsymbol{\gamma}_j, \alpha_j, \mathbf{b}_{j,0}, \mathbf{b}_{j,1}, \sigma_j^2) \\ \times \pi(\alpha_j \mid \sigma_j^2) \pi(\mathbf{b}_{j,0}, \mathbf{b}_{j,1} \mid \sigma_j^2, \boldsymbol{\gamma}_j) \pi(\sigma_j^2 \mid \boldsymbol{\gamma}_j) \pi(\boldsymbol{\gamma}_j) d\alpha_j d\mathbf{b}_{j,0} d\mathbf{b}_{j,1} d\sigma_j^2 d\boldsymbol{\gamma}_j. \end{aligned} \tag{5}$$

Intuitively, the marginal likelihood takes the expectation of the unknown parameters in the likelihood function with respect to the prior distributions. A function of the data and prior parameters only, it accounts for parameter estimation and model selection uncertainties, separating the tree growth from factor model estimations. We describe in Proposition 1 (proved in Appendix II.) the marginal likelihood.

Proposition 1. *Stacking all asset returns in the node in a vector* $\mathbf{R}$*, the marginal likelihood in (5) has a closed form:*

$$p(\mathbf{R} \mid \mathbf{Z}, \mathbf{F}) = \sum_{I_m} \pi(\boldsymbol{\gamma}_{j,\mathbf{f}} = I_m) p(\mathbf{R} \mid \boldsymbol{\gamma}_j, \mathbf{Z}, \mathbf{F}), \tag{6}$$

$$\textit{where} \quad p(\mathbf{R} \mid \boldsymbol{\gamma}_j, \mathbf{Z}, \mathbf{F}) = \frac{1}{(2\pi)^{N/2}} \sqrt{\frac{|\boldsymbol{\Lambda}_{0|\boldsymbol{\gamma}_j}|}{|\boldsymbol{\Lambda}_N|}} \frac{v_0^{S_0}}{v_N^{S_N}} \frac{\Gamma(S_N)}{\Gamma(S_0)},$$

$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$ is the gamma function, and $\pi(\boldsymbol{\gamma}_{j,\mathbf{f}} = I_m) = \prod_{k=1}^K w_k^{\gamma_{j,\mathbf{f},k}} (1-w_k)^{(1-\gamma_{j,\mathbf{f},k})}$ is the prior probability of the model as shown in (4). I_m is a length K vector with all elements 0 or 1, and $\{I_m\}_{m=1}^{2^K}$ enumerates all possible values that $\boldsymbol{\gamma}_{j,\mathbf{f}}$ can take. Note that $\boldsymbol{\gamma}_j$ is uniquely defined once $\boldsymbol{\gamma}_{j,\mathbf{f}}$ is given. Furthermore,

$$\begin{aligned} \boldsymbol{\Lambda}_N &= \mathcal{W}^\mathsf{T} \mathcal{W} + \boldsymbol{\Lambda}_{0|\boldsymbol{\gamma}_j}, & \boldsymbol{\mu}_N &= \boldsymbol{\Lambda}_N^{-1} \left(\mathcal{W}^\mathsf{T} \mathbf{r} \right), \\ v_N &= v_0 + \frac{1}{2} \left(\mathbf{R}^\mathsf{T} \mathbf{R} - \boldsymbol{\mu}_N^\mathsf{T} \boldsymbol{\Lambda}_N \boldsymbol{\mu}_N \right), & S_N &= S_0 + \frac{N}{2}, \end{aligned}$$

where N is total number of data observations, $\mathcal{W} = [1, \mathbf{F}, \mathbf{F} \bullet \mathbf{Z}]$, and $\boldsymbol{\Lambda}_{0|\boldsymbol{\gamma}_j}$ is a diagonal matrix with s-th element being ξ_1^{-2} if $\gamma_{j,s} = 1$, or ξ_0^{-2} if $\gamma_{j,s} = 0$ and 0 elsewhere.

To simplify the notation, we now denote the marginal likelihood evaluation of (6)

using all observations in node $\mathcal{A}_n$, $n \in \{0, 1, 2, \dots\}$ by $p(\mathcal{A}_n)$.

Split and stopping criteria. We collect all split rule candidates in $\mathcal{C} = \{\tilde{c}_j\}$, including candidates that split each variable at all possible cutpoint values. Recall that each split rule candidate $\tilde{c}_j$ partitions the current node into two child leaf nodes (Figure 2) before we fit the factor model at each potential leaf cluster separately. Because $\epsilon_{i,t}$ are independent, the two child nodes are independent, and the joint marginal likelihood of the entire data is the product of the marginal likelihoods of the two child nodes,

$$l(c_j) = p(\mathbf{R}^L \mid \mathbf{Z}^L, \mathbf{F}^L) \times p(\mathbf{R}^R \mid \mathbf{Z}^R, \mathbf{F}^R) = p(\mathcal{A}_1) \times p(\mathcal{A}_2). \tag{7}$$

The superscript L and R represent data observations partitioned to the left and right child nodes by the split rule candidate $\tilde{c}_j$. Each split rule candidate partitions the data to $\mathcal{A}_1$ and $\mathcal{A}_2$ differently, and the value of (7) varies across split rule candidates, serving as a natural measurement of model fitness and quality of the candidate with larger values better.

Furthermore, besides all the split rule candidates, we consider the option to stop splitting at the current node definitively, the *null cutpoint* option, the marginal likelihood of which is defined as:

$$l(\emptyset) = |\mathcal{C}| \left(\frac{(1+d)^{\tilde{a}_2}}{\tilde{a}_1} - 1 \right) p(\mathbf{R} \mid \mathbf{Z}, \mathbf{F}) = |\mathcal{C}| \left(\frac{(1+d)^{\tilde{a}_2}}{\tilde{a}_1} - 1 \right) p(\mathcal{A}_0), \tag{8}$$

where $|\mathcal{C}|$ is the number of all split rule candidates, d is the depth of the current node (root node has depth 1), and $\tilde{a}_1$ and $\tilde{a}_2$ are two hyper-parameters. The marginal likelihood in (8) is evaluated on *all* data in the root node, since all the data belong to the same current node if we stop splitting further. We follow the XBART framework (He et al., 2019; He and Hahn, 2021) to set the early stop option to match the regularization tree prior used in BART (Chipman, George, and McCulloch, 2010) that chooses $\tilde{a}_1 = 0.95$ and $\tilde{a}_2 = 2$. The additional weight on the null cutpoint criterion $|\mathcal{C}| \left(\frac{(1+d)^{\tilde{a}_2}}{\tilde{a}_1} - 1 \right)$ is chosen as regularization of the tree. As the tree grows, d increases, and the weight (penalty) increases exponentially. If the benefit of splitting does not outweigh the growth penalty on tree size, the tree stops growing to avoid overfitting.

Once all marginal likelihoods have been calculated for all split rule candidates and the null cutpoint, assuming the equal prior probability of each split rule candidate, the split criterion can be derived following the Bayes rule:

$$L(c_j) = \frac{l(c_j)}{\sum_{j'} l(c_{j'}) + l(\emptyset)}, \quad L(\emptyset) = \frac{l(\emptyset)}{\sum_{j'} l(c_{j'}) + l(\emptyset)}. \tag{9}$$

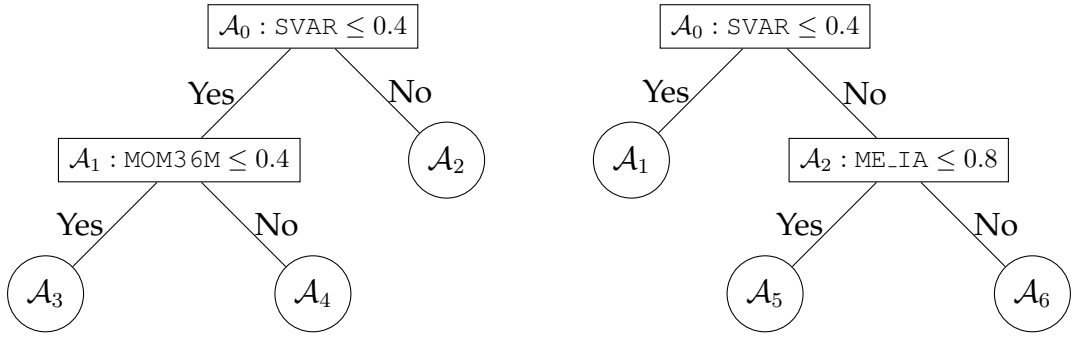
The split criterion in (9) is essentially the posterior probability of a split candidate with the prior probability of each candidate proportional to 1. We choose the one that maximizes (9) as the first split rule. This is a maximum posterior (MAP) estimator since it maximizes the posterior probability. Once a split rule is selected, we split the current node and yield two child leaf nodes. However, if the null cutpoint is selected, the current node will not split further.

2.3.2 Second Split

If the first split is not the null cutpoint, the second split can happen at either the left or right child node of the root node. Therefore all split candidates of both nodes should be considered, and the split rule candidates expand to $\mathcal{C} = \{c_k^{\mathcal{A}_1}\} \cup \{c_k^{\mathcal{A}_2}\}$, where the superscript denotes which node to split. Figure 3 illustrates two split rule candidates, for example, at either side of the root.

Figure 3: **Illustration of the Second Split Candidates**

This figure illustrates two example candidates for the second split. Note that no matter which leaf node to split, the second (j-th) iteration of the tree algorithm has three ($j+1$) leaf nodes.



(a) If splitting node $\mathcal{A}_1$ at MOM3 6M.

(b) If splitting node $\mathcal{A}_2$ at ME_IA.

No matter which one is chosen, one leaf node splits into two new leaf nodes for

the next iteration. The joint marginal likelihood corresponding to Eq. (7) is:

$$l(c_j^{\mathcal{A}_1}) = p(\mathcal{A}_3) \times p(\mathcal{A}_4) \times p(\mathcal{A}_2), \quad l(c_j^{\mathcal{A}_2}) \quad = p(\mathcal{A}_1) \times p(\mathcal{A}_5) \times p(\mathcal{A}_6), \tag{10}$$

which is essentially the product of the marginal likelihood evaluated on *all* resulting leaves. The marginal likelihood of a null cutpoint is defined similarly:

$$\begin{aligned} l(\emptyset^{\mathcal{A}_1}) &= |\mathcal{C}| \left(\frac{(1+d)^{\tilde{a}_2}}{\tilde{a}_1} - 1 \right) p(\mathcal{A}_1) \times p(\mathcal{A}_2) \\ l(\emptyset^{\mathcal{A}_2}) &= |\mathcal{C}| \left(\frac{(1+d)^{\tilde{a}_2}}{\tilde{a}_1} - 1 \right) p(\mathcal{A}_2) \times p(\mathcal{A}_1), \end{aligned}$$

where the null cutpoints $\emptyset^{\mathcal{A}_1}$ and $\emptyset^{\mathcal{A}_2}$ denote stop splitting the node $\mathcal{A}_1$ or $\mathcal{A}_2$, respectively. Note that $d = 2$ for the second split, since $\mathcal{A}_1$ or $\mathcal{A}_2$ has depth 2.

Similarly, for all candidates in $\mathcal{C} = \{c_j^{\mathcal{A}_1}\} \cup \{c_j^{\mathcal{A}_2}\}$ and two nulls, we have:

$$\begin{aligned} W &= \sum_{j'} l(c_{j'}^{\mathcal{A}_1}) + \sum_{j'} l(c_{j'}^{\mathcal{A}_2}) + l(\emptyset^{\mathcal{A}_1}) + l(\emptyset^{\mathcal{A}_2}) \\ L(c_j^{\mathcal{A}_1}) &= \frac{l(c_j^{\mathcal{A}_1})}{W}, L(\emptyset^{\mathcal{A}_1}) = \frac{l(\emptyset^{\mathcal{A}_1})}{W}, L(c_j^{\mathcal{A}_2}) = \frac{l(c_j^{\mathcal{A}_2})}{W}, L(\emptyset^{\mathcal{A}_2}) = \frac{l(\emptyset^{\mathcal{A}_2})}{W}. \end{aligned} \tag{11}$$

The second split is the one maximizing (11). We emphasize that the split criterion is defined *globally* — (10) is defined on all resulting leaf nodes and all data. Thus, the second split maximizes the global marginal likelihood to avoid overfitting. This global split criterion avoids the myopic local split criterion used in standard machine learning recursive algorithms. The node no longer splits when the null cutpoint is chosen for a specific node, say $\mathcal{A}_1$.

2.3.3 Further Splits

The clustering algorithm proceeds iteratively by splitting leaf nodes based on marginal likelihood improvement. All existing leaf nodes are considered for each split, and the split rule candidates are evaluated. The global split criterion is defined on all resulting leaves, and the candidates that maximize the split criterion are chosen. The tree-growing process terminates once all leaf nodes have chosen the early stop option or some pre-specified stopping conditions are met. The tree creates J leaf clusters if

the tree terminates splitting after J-1 iterations. It is worth repeating that by using marginal likelihood, we avoid intermediate model estimation errors. Yet, the choice of split rules, or tree growth, is still determined by Bayesian variable selection in each resulting cluster. Moreover, the joint marginal likelihood has a natural economic interpretation, whereas a global split criterion under a non-Bayesian framework has to be defined ad hoc.

2.3.4 Post-Clustering Inference

Once the tree growth is complete, we refit the model using (2) for each cluster, and obtain posterior samples of all model parameters by Markov chain Monte Carlo (MCMC) approach, using the Gibbs sampler for inference. Expressing the stacked data in matrix form, we have

$$\mathbf{R} = \alpha_j + \mathbf{b}_{j,0}^{\mathsf{T}}\mathbf{F} + \mathbf{b}_{j,1}^{\mathsf{T}}\left(\mathbf{F} \bullet \mathbf{Z}\right) + \boldsymbol{\epsilon} = \mathcal{W}^{\mathsf{T}}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{12}$$

where $\mathcal{W}^{\mathsf{T}} = [\mathbf{1}, \mathbf{F}, \mathbf{F} \bullet \mathbf{Z}]$ and $\boldsymbol{\beta} = [\alpha_j^{\mathsf{T}}, \mathbf{b}_{j,0}^{\mathsf{T}}, \mathbf{b}_{j,1}^{\mathsf{T}}]^{\mathsf{T}}$ to simplify notations. The full conditionals are given as follows,

1. Update $\boldsymbol{\beta} = \{\beta_k\}_{k=0,\cdots,K+MK}$. Note that this step updates the regression coefficients of all variables. For simplicity, we represent them using β_k rather than distinguishing $[\alpha_j, \mathbf{b}_{j,0}, \mathbf{b}_{j,1}]$. For $k = 0, 2, \cdots, K + MK$,

$$\beta_k \mid \boldsymbol{\beta}_{-k}, \sigma_j^2, \boldsymbol{\gamma}_j, \mathbf{R}, \mathcal{W} \sim \mathcal{N}\left(\tilde{\beta}_k, \frac{\sigma_j^2}{\mu_k}\right),$$

   where $\mu_k = \mathbf{w}_k^{\mathsf{T}}\mathbf{w}_k + \frac{1}{\xi_{\gamma_{j,k}}}$ and $\tilde{\beta}_k = \mu_k^{-1}\mathbf{w}_k^{\mathsf{T}}(\mathbf{R} - \mathcal{W}_{-k}\boldsymbol{\beta}_{-k})$. Note that $\mathbf{w}_k$ and $\mathcal{W}_{-k}$ denote the k-th column and all columns except the k-th of matrix $\mathcal{W}$ respectively, and $\xi_{\gamma_{j,0}} = \xi_1$ for the intercept $\beta_0 = a_j$ always in the "slab".

2. Update $\boldsymbol{\gamma}_{j,\mathbf{f},k}$, for $k = 1, 2, \cdots, K$,

$$\gamma_{j,\mathbf{f},k} \mid \boldsymbol{\beta}, \sigma_j^2, \mathbf{R}, \mathcal{W} \sim \textsc{Ber}\left(\frac{\xi_1^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma_j^2\xi_1}\beta_k^2\right)w_k}{\xi_1^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma_j^2\xi_1}\beta_k^2\right)w_k + \xi_0^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma_j^2\xi_0}\beta_k^2\right)(1-w_k)}\right).$$

Then update $\boldsymbol{\gamma}_{j,\mathbf{f}\bullet\mathbf{z}}$, copy the corresponding values in $\boldsymbol{\gamma}_{j,\mathbf{f}}$ for the main effect of $\mathbf{f}_t$, and assemble $\boldsymbol{\gamma}_j = (\boldsymbol{\gamma}_{j,\mathbf{f}}, \boldsymbol{\gamma}_{j,\mathbf{f}\bullet\mathbf{z}})$.

3. Update σ_j^2

$$\sigma_j^2 \mid \boldsymbol{\beta}, \boldsymbol{\gamma}_j, \mathbf{R}, \mathcal{W} \sim IG(S^*, v^*),$$

where $S^* = \frac{1}{2}(N + K + KM + S_0)$, $v^* = \frac{1}{2}\left(||\mathbf{R} - \mathcal{W}\boldsymbol{\beta}||_2^2 + \sum_{k=0}^{K+MK} \frac{\beta_k^2}{\xi_{\gamma_{j,k}}} + v_0\right)$, and N is number of data observations in the leaf node.

3 Data on U.S. Equities

Assets. To illustrate the performance of BCM, we apply it to splitting and pricing the whole universe of U.S. individual equities at the monthly frequency. Our sample spans from January 1972 to December 2021. We apply the standard filters (see, e.g., Fama and French, 2015) to (1) only include stocks listed on NYSE, AMEX, or NASDAQ for more than one year, (2) use observations of firms with CRSP share codes of 10 and 11, and (3) exclude stocks with negative book equity or negative lagged market equity. We use the first 35-year subsample for model estimation, and the recent 15 years for out-of-sample tests. In the training sample, the average and median number of stocks are 4,759 and 4,733, respectively, while in the testing sample, these numbers are 3,742 and 3,633. Note that BCM allows the panel data to be unbalanced.

Factors. Although BCM does not impose restrictions on the candidate factors (both tradable and non-tradable) or variables to select from in estimating the heterogeneous models in each leaf cluster, we consider a collection of 20 tradable factors for easy interpretation and computation. These factors are widely adopted and prominent in the asset pricing literature and suffice for illustrating the power of BCM. Alternative factors can be easily added to BCM.

First, we include factors from two popular factor models, the five factors from Fama and French (2015) of excess market factor (MKTRF), small-minus-big (SMB), high-minus-low (HML), robust-minus-weak (RMW), and conservative-minus-aggressive (CMA), and the additional q-factors (Hou, Xue, and Zhang, 2015; Hou, Mo, Xue, and Zhang, 2021) of investment (IA), profitability (ROE), and expected growth (REG). We

also collect popular factors from the investment profession: betting-against-beta (BAB, Frazzini and Pedersen, 2014), quality-minus-junk (QMJ, Asness, Frazzini, and Pedersen, 2019), and HML Devil (HMLM, Asness and Frazzini, 2013).

Finally, we include commonly used factors in the literature: market beta (BETA, Fama and MacBeth, 1973), momentum (UMD, Carhart, 1997), long-term reversal (LTR, De Bondt and Thaler, 1985), short-term reversal (STR, Jegadeesh and Titman, 1993), liquidity (LIQ, Pástor and Stambaugh, 2003), idiosyncratic volatility (IVOL, Ali, Hwang, and Trombley, 2003), intermediary asset pricing (IMD, He, Kelly, and Manela, 2017), earning surprise (SUE, Rendleman Jr, Jones, and Latane, 1982), and net equity issues (NI, Loughran and Ritter, 1995).

Characteristics. We use 20 firm characteristics with monthly observations for each asset: market equity (ME), market beta (BETA), bid-ask spread (BAS), stock return variance (SVAR), book-to-market (BM), earnings-to-price (EP), cash flows-to-price (CFP), sales-to-price (SP), asset growth (AGR), net equity issuance (NI), accounting accruals (ACC), operating profitability (OP), return on equity (ROE), momentum (MOM), short-term reversal (STR), seasonality (SEAS), advertisement-to-market (ADM), R&D-to-market (RDM), unexpected quarterly earning (SUE), and earning- announcement abnormal return (ABR).[8] We rank all observations of each characteristic in each month cross-sectionally and standardize the rank scale in the range of [0, 1], with their cross-sectional median set to 0.5. Missing values are then imputed to be 0.5.[9] We use quintiles as split value candidates for each characteristic to mimic conventional 5×1 sorting in asset pricing. For 20 characteristics, there are 80 (4×20) splitting candidates.

Macroeconomic variables. Following Welch and Goyal (2008), we adopt ten macro variables for detecting time-series breaks and regimes relevant to asset pricing. These variables include the 3-month treasury bill rate, inflation, term spread, default spread, dividend yield, earnings-to-price, market volatility, net equity issues, leverage, and liquidity. We map the macroeconomic variables into the $[0, 1]$ range based on their

[8] The characteristics cover both firm fundamentals and market signals, span all categories commonly used in the literature, and have been mostly introduced before the test sample period.

[9] In earlier drafts, we standardized the values directly to $[-1, 1]$ and obtained similar results. We report results using rank-based standardization because of its greater robustness in general.

empirical percentile values in a rolling 10-year window. For example, if the inflation variable exceeds 0.7, meaning the current inflation level is higher than 70% of observations in the past ten years. The standardization enables us to compare the performance of each predictor on the same variation scale without look-ahead biases.

4 Empirical Application to Asset Pricing

This section presents our empirical findings applying the baseline BCM to modeling asset heterogeneity and uncommon factors in the cross section of stock returns; Section 5 extends the model to include structural breaks and macro-instrumented regimes to allow for grouped heterogeneity in both the cross section and time series. Throughout the discussion, we set $w_i = 0.5$ for agnostic investors to indicate their equal prior probability to select any factor, and $w_i = 0.1$ for skeptic investors to reflect their prior belief that most factors are irrelevant.

4.1 Data-Driven and Economically Guided Asset Clusters

BCM clusters the unbalanced panel of individual stock returns by iteratively splitting the entire sample using firm characteristics, based on marginal likelihood improvement as the global split criteria. Therefore, it constitutes an economically guided panel tree (Cong et al., 2022). The upper panel of Figure 4 shows the evaluations based on the split criterion of all candidates for the first split.[10] The heat map depicts the resulting marginal likelihood for all 80 possible split candidates, with the value obtained with an SVAR split at a threshold of 0.4 (S1), followed by SVAR at 0.6, ME at 0.4, and EP at 0.4. The importance of volatility, size, and value in splitting the cross section of individual equity returns is consistent with the broad finance literature.[11] Moreover, all split rule candidates have a higher marginal likelihood than the root, indicating the presence of asset heterogeneity and the benefit of clustering.

The first split yields two leaf nodes; the second can occur on either side (see Figure 2). The heat map in the lower panels displays the evaluations of split criteria for all

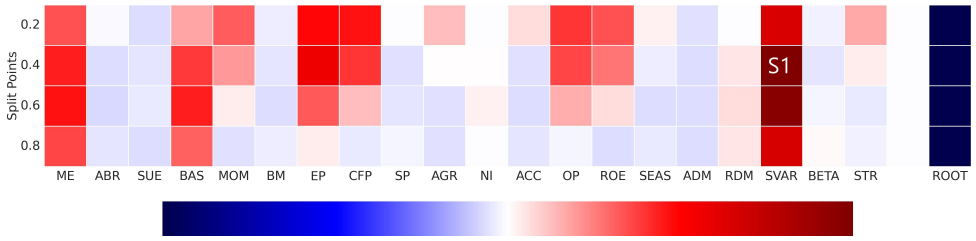
[10] Although finer grids of cutpoints are feasible, they do not alter the main results. It suffices to consider quintile cutpoints of each characteristic, as in the conventional 5×1 sorting.

[11] For example, Pontiff (2006) shows that idiosyncratic volatility reflects holding costs and is thus related to mispricing.

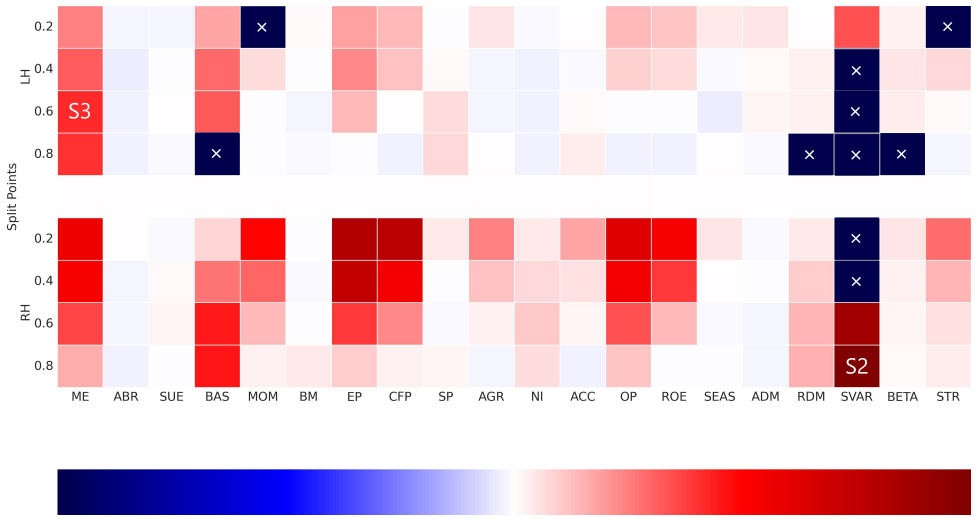
Figure 4: First Two Splits in the Cross Section

The heat map in the upper panel displays the search for the first split's cutpoint, aiming for the highest marginal likelihood value. Twenty characteristics and their cross-sectional quintiles are considered as candidates for cutpoints. The heat map in the lower panel shows the search for the second split's cutpoint, also aiming for the highest marginal likelihood value. The tree algorithm must compute 80 $(K+1)$ candidates for iterative splitting.

(a) First Split



(b) Second Split



candidate splits. Each panel represents a split at the left and right leaves, resulting in 80×2 split candidates. Certain candidates may generate small clusters in certain months due to limited data on specific interactions of correlated characteristics. We set a minimum leaf size of 10 stock returns per month to ensure an adequate number of observations. Candidates without child leave meeting this requirement are discarded (indicated by dark cells with a white cross X). The optimal second split, S2, is again SVAR (at 0.8) and is implemented on the right leaf node after the first split (N1 S1 for SVAR ≤ 0.4).

For brevity, we leave out demonstrations for subsequent splits. K splits yield $K+1$ leaves, and the algorithm considers $80 \times (K+1)$ possible candidates for iterative

splitting. Figure A.1 (in the appendix) displays the improvement in (log) marginal likelihood for each split compared to the root node, accounting for the normalizing constants. The initial split significantly increases the marginal likelihood, as do many subsequent splits, corroborating the presence of grouped heterogeneity. As the benefit of additional splits eventually diminishes, tree growth stops after 20 splits due to the early stopping in (8).

The BCM tree (Figure 5) comprises 21 terminal leaves, representing 21 clusters managed through various asset fundamentals and market signals. The sequence of split rules from the root to the leaves based on the lagged firm characteristics determines the cluster membership of an asset, which may change over time due to changes in asset characteristics. The average monthly return observations for the terminal leaf nodes are reported. The resulting leaf clusters appear balanced, neither too large nor too small.

The leaf nodes fall into four mega clusters based on the first three splits in the top two layers of the tree: (i) low-variance and small-cap (SVAR ≤ 0.4 and ME ≤ 0.6), (ii) low-variance and large-cap (SVAR ≤ 0.4 and ME > 0.6), (iii) medium-variance ($0.4 <$ SVAR ≤ 0.8), and (iv) high volatility (SVAR > 0.8).

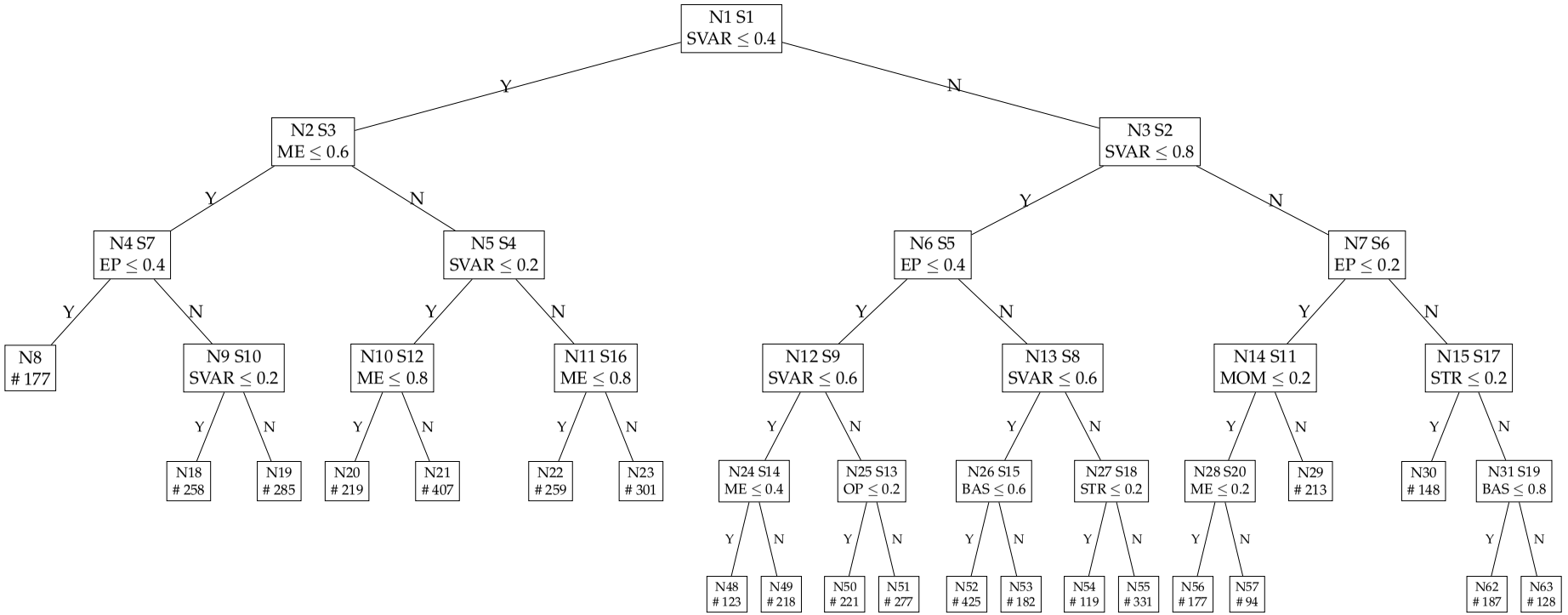
In the third layer, earnings-to-price (EP) is the splitting candidate at three out of four nodes (N4, N6, N7). Other characteristics used in further splits include one-year momentum (MOM), short-term reversal (STR), operating profitability (OP), and bid-ask spread (BAS). Such splitting of the cross section results in characteristics-based clusters of assets that may have different factor structures, reflecting asset heterogeneity, and are used as test assets on the left-hand side in (1) in what follows.

4.2 Common and Uncommon Factors

Factor selection and common factors. The posterior mean of the latent variables $\gamma_{j,\mathbf{f}}$ has a natural interpretation as the posterior probability that a factor is selected. Table 1 presents the factor selection probabilities in each leaf cluster, which are computed based on 3,500 MCMC samples with the first 1,500 burn-in draws discarded. The IDs of leaf clusters correspond to the terminal leaves of the tree in Figure 5. In parentheses, we also report the Bayesian estimates of the betas for various factors. Given that betas

Figure 5: Tree Cluster

Using monthly data from 1972 to 2006, the tree divides individual stock returns based on cross-sectional standardization of firm characteristic ranks to [0,1]. The resulting terminal leaves correspond to clusters identified by the firm characteristics. Each node, including leaves and intermediate nodes, has an ID indicated by $N\#$, and the order of the splits is denoted by $S\#$. For instance, $S3$ refers to the third split.



in BCM are time-varying as functions of firm characteristics, we first compute their time-series average for individual stock returns in each leaf cluster and then present their equal-weighted averages across individual betas to measure cluster exposures to the factors.

MKTRF is selected with a 100% probability in most leaf clusters, except for N54 (0%), N56 (22.6%), and N63 (88.7%). This finding confirms the market factor's role as a universal pricing factor, aligning with expectations from CAPM. The size factor (SMB) is also crucial, selected with a probability exceeding 0.7 in 13 out of 21 clusters. STR and BETA are similarly prominent, with selection probabilities greater than 0.7 in 12 and 10 clusters, respectively. However, the remaining factors demonstrate limited usefulness, appearing in only a few clusters. For instance, the profitability factors (RMW and ROE) are selected with probabilities above 0.7 in just 3 clusters, while the investment factors (CMA and IA) exhibit low selection probabilities across most leaf clusters.

Note that Harvey and Liu (2021) similarly identify MKTRF and SMB as the most important factors. As in that study, the fact that MKTRF and SME have consistent significance and signs across clusters here is nontrivial. For example, while the market factor has theoretical underpinnings and was the first risk factor tested (Jensen et al., 1972), various studies have shown that it becomes insignificant or even reverses its sign when combined with other factors, casting doubts on its empirical validity as a risk factor. Nevertheless, it is widely used in risk adjustments and cost-of-capital calculations, probably due to its inherent intuitiveness and theoretical appeal.

N56 is a special cluster with selection probabilities below 50% for almost all factors, and negligible betas on those factors. It represents assets with the highest volatility (SVAR > 0.8), poor earnings (EP ≤ 0.2), past loser characteristics (MOM ≤ 0.2), and small size (ME ≤ 0.2). These factors are typically associated with the most challenging stocks in empirical asset pricing (see, e.g., Fama and French, 1996, 2015). In comparison, N57, which shares the same parent node as N56, only differs slightly in size but has a 100% selection probability for MKTRF, SMB, and UMD. BCM informs us that N56 assets are the least "priceable."

Table 1: Factor Selection Probability and Beta Estimates

This table presents the probability of factor selection for each leaf cluster, computed from 2,000 MCMC samples, with the leaf cluster IDs matching those of the tree in Figure 5. The Bayesian estimate of the factor betas, which includes the posterior average for spike and slab states, is also presented in parentheses.

Factors	N8	N18	N19	N20	N21	N22	N23	N48	N49	N50	N51
MKTRF	100 (0.406)	100 (0.342)	100 (0.680)	100 (0.881)	100 (0.925)	100 (0.962)	100 (1.088)	100 (0.562)	100 (1.008)	100 (0.703)	100 (0.891)
SMB	0.0 (0.000)	0.0 (0.000)	100 (0.706)	100 (0.756)	100 (0.263)	100 (0.742)	100 (0.470)	0.0 (0.000)	75.5 (0.686)	0.0 (0.000)	100 (0.790)
HML	52.2 (0.029)	66.0 (0.128)	100 (0.245)	100 (0.410)	100 (0.802)	0.6 (0.028)	66.0 (0.640)	30.5 (-0.152)	66.0 (-0.191)	22.3 (-0.030)	45.4 (0.017)
RMW	21.8 (0.049)	33.7 (0.049)	100 (0.232)	63.4 (0.140)	21.8 (0.039)	5.1 (0.013)	100 (0.205)	19.0 (0.035)	36.0 (-0.005)	5.2 (-0.008)	2.0 (0.001)
CMA	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	3.1 (0.001)	0.0 (0.000)	69.3 (0.291)	0.0 (0.000)	23.5 (0.103)	0.0 (0.000)
UMD	20.5 (-0.010)	100 (-0.074)	95.8 (-0.142)	100 (-0.092)	0.0 (0.000)	100 (-0.141)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	76.1 (-0.125)	81.1 (-0.068)
HMLM	4.1 (-0.003)	35.1 (0.034)	0.0 (0.000)	0.0 (0.000)	100 (0.238)	19.0 (0.019)	100 (0.187)	15.2 (0.061)	100 (0.086)	0.0 (0.000)	0.0 (0.000)
QMJ	63.3 (0.022)	1.1 (0.001)	59.1 (-0.017)	12.0 (0.019)	0.0 (0.000)	0.0 (0.000)	0.4 (0.000)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)
BAB	88.9 (0.562)	100 (0.573)	67.9 (0.239)	0.0 (0.000)	100 (0.126)	83.2 (0.226)	3.4 (0.001)	51.4 (0.300)	78.4 (0.144)	5.1 (0.021)	20.8 (0.036)
LIQ	0.7 (0.000)	6.9 (-0.001)	2.3 (-0.001)	4.1 (-0.001)	6.9 (0.000)	9.4 (-0.005)	100 (-0.048)	0.6 (0.000)	1.5 (0.000)	0.3 (0.000)	2.3 (-0.001)
STR	14.4 (0.008)	100 (0.065)	100 (0.122)	2.9 (0.000)	100 (0.037)	0.0 (0.000)	100 (0.015)	100 (0.143)	100 (0.080)	15.1 (0.040)	100 (0.186)
LTR	91.0 (0.077)	91.0 (0.101)	0.5 (0.000)	1.3 (-0.001)	9.2 (-0.001)	1.0 (-0.001)	2.0 (0.002)	44.3 (0.108)	24.9 (0.076)	13.4 (0.051)	1.9 (-0.001)
REG	0.0 (0.000)	0.0 (0.000)	44.0 (-0.004)	0.0 (0.000)	27.1 (-0.028)	0.0 (0.000)	87.7 (-0.147)	3.5 (-0.002)	17.8 (-0.021)	0.0 (0.000)	0.0 (0.000)
IVOL	100 (0.286)	100 (0.159)	100 (-0.029)	100 (-0.160)	8.8 (-0.009)	18.4 (-0.039)	6.3 (-0.001)	100 (0.490)	29.9 (0.082)	99.9 (0.646)	100 (0.135)
SUE	81.4 (-0.107)	100 (-0.043)	6.7 (-0.004)	0.0 (0.000)	43.3 (0.015)	56.4 (0.048)	55.4 (0.020)	0.0 (0.000)	10.0 (0.001)	5.2 (-0.005)	95.0 (0.083)
IA	40.6 (0.069)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	49.6 (0.015)	0.0 (0.000)	1.2 (0.001)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)
ROE	20.5 (-0.067)	4.7 (0.001)	9.4 (-0.008)	0.0 (0.000)	100 (0.302)	3.3 (0.010)	100 (0.301)	0.0 (0.000)	0.0 (0.000)	15.5 (-0.075)	100 (-0.324)
IMD	1.0 (0.000)	0.2 (0.000)	5.2 (0.000)	9.4 (0.002)	2.3 (0.001)	0.0 (0.000)	0.0 (0.000)	29.0 (0.020)	0.0 (0.000)	1.5 (0.000)	0.0 (0.000)
NI	4.6 (-0.005)	67.3 (-0.059)	2.2 (0.000)	1.2 (0.000)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	1.8 (0.001)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)
BETA	100 (0.154)	100 (0.145)	100 (0.118)	1.4 (-0.001)	100 (-0.019)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	78.5 (0.080)	4.8 (0.006)	100 (0.098)

Factors	N52	N53	N54	N55	N56	N57	N29	N30	N62	N63
MKTRF	100 (0.944)	100 (0.891)	0.0 (0.000)	100 (0.718)	22.6 (0.083)	100 (1.153)	100 (0.704)	100 (0.833)	100 (0.927)	88.7 (0.695)
SMB	100 (0.803)	100 (0.754)	75.1 (0.568)	0.0 (0.000)	10.9 (0.096)	100 (1.580)	41.2 (0.307)	50.2 (0.342)	100 (0.751)	100 (0.836)
HML	100 (0.261)	66.0 (0.150)	2.6 (0.006)	10.2 (0.028)	21.1 (-0.003)	11.7 (0.007)	4.2 (-0.046)	67.6 (0.102)	66.0 (0.155)	21.8 (0.012)
RMW	100 (0.267)	15.0 (0.014)	21.8 (0.068)	63.4 (0.123)	16.5 (-0.075)	0.2 (-0.004)	7.6 (-0.008)	5.3 (-0.005)	3.2 (-0.003)	6.4 (0.004)
CMA	0.5 (0.000)	0.0 (0.000)	29.7 (0.125)	1.1 (0.000)	0.0 (0.000)	0.0 (0.000)	7.6 (0.031)	11.4 (0.027)	0.0 (0.000)	0.0 (0.000)
UMD	100 (-0.143)	61.8 (-0.106)	1.2 (-0.001)	100 (-0.128)	30.3 (-0.201)	100 (-1.031)	0.0 (0.000)	42.8 (-0.053)	31.1 (-0.018)	0.0 (0.000)
HMLM	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	2.0 (0.001)	9.6 (0.081)	0.0 (0.000)	10.3 (0.031)	0.0 (0.000)	0.0 (0.000)	7.4 (0.006)
QMJ	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	6.7 (-0.116)	14.5 (-0.230)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	100 (-0.601)
BAB	46.5 (0.070)	98.8 (0.222)	0.0 (0.000)	100 (0.449)	5.0 (0.021)	0.0 (0.000)	0.4 (0.001)	11.4 (0.028)	4.5 (0.002)	3.9 (0.004)
LIQ	6.9 (-0.004)	0.7 (0.000)	6.9 (0.007)	5.4 (-0.001)	21.7 (-0.004)	8.0 (-0.005)	6.9 (-0.007)	4.1 (-0.001)	3.4 (-0.003)	6.9 (0.004)
STR	100 (0.139)	39.8 (0.043)	100 (0.600)	100 (0.077)	34.6 (0.201)	0.0 (0.000)	0.0 (0.000)	100 (0.541)	70.0 (0.046)	100 (-0.077)
LTR	2.2 (0.000)	1.9 (-0.001)	13.3 (0.025)	91.0 (0.201)	1.8 (0.016)	23.5 (0.000)	64.0 (0.314)	43.8 (0.177)	0.4 (0.000)	6.0 (0.003)
REG	2.0 (-0.001)	0.0 (0.000)	3.7 (0.000)	6.2 (-0.007)	43.3 (-0.059)	0.2 (-0.002)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)
IVOL	0.5 (0.000)	1.8 (0.001)	100 (0.237)	100 (0.411)	66.0 (0.581)	1.8 (0.012)	100 (0.610)	100 (0.636)	100 (0.368)	100 (0.247)
SUE	0.0 (0.000)	39.0 (-0.045)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	2.0 (-0.007)	0.4 (0.000)	0.0 (0.000)	44.2 (-0.042)	0.0 (0.000)
IA	100 (0.000)	4.9 (0.000)	40.6 (-0.164)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	0.0 (0.000)	45.0 (0.081)	0.0 (0.000)
ROE	0.6 (0.000)	0.0 (0.000)	0.0 (0.000)	0.4 (0.000)	2.7 (-0.032)	0.0 (0.000)	47.5 (-0.330)	9.7 (-0.033)	0.0 (0.000)	0.0 (0.000)
IMD	2.8 (0.001)	0.0 (0.000)	100 (0.405)	57.8 (-0.033)	1.2 (0.002)	0.0 (0.000)	5.3 (0.003)	0.0 (0.000)	0.0 (0.000)	13.0 (0.035)
NI	0.8 (0.000)	2.6 (-0.001)	6.0 (-0.004)	0.2 (0.000)	13.1 (0.013)	0.0 (0.000)	0.0 (0.000)	81.3 (-0.114)	81.3 (-0.051)	25.3 (-0.029)
BETA	100 (0.100)	85.5 (0.087)	100 (0.320)	100 (0.188)	59.8 (0.432)	0.0 (0.000)	22.6 (0.049)	0.0 (0.000)	0.0 (0.000)	15.1 (0.026)

Uncommon factors. While MKTRF and SMB are common factors, many other factors are selected with high probability only in a subset of clusters, and are, therefore, uncommon and asset-specific. To understand which factors price what assets, we further summarize the factor selection results in Table 1 using different probability thresholds in Table 2. Specifically, we consider three thresholds of 0.95, 0.9, and 0.7, to categorize factors based on their selection probabilities. This way, we clarify which factors are common across clusters and which are uncommon and specific for which clusters.

Table 2: Factor Selection Evaluation by Leaf Clusters

This summary table presents the results from Table 1 under different selection probability thresholds. The selected factors are listed according to their respective probability values, with rows representing probability thresholds of > 0.95, > 0.9, and > 0.7. Four panels display leaf cluster positions for the top three tree layers. We present in-sample (1972-2006) and out-of-sample (2007-2021) alphas (in %) for cluster-specific fitted factor models.

Prob.	N8	N18	N19		N48	N49	N50	N51	N52	N53	N54	N55
	Panel A: I(SVAR $\leq$ 0.4)I(ME $\leq$ 0.6)				Panel C: I(SVAR $>$ 0.4)I(SVAR $\leq$ 0.8)							
> 0.95	MKTRF IVOL BETA	MKTRF IVOL BETA BAB SUE UMD STR	MKTRF IVOL BETA STR HML RMW SMB UMD		MKTRF STR IVOL	MKTRF STR HMLM	MKTRF IVOL	MKTRF IVOL ROE STR BETA SMB	MKTRF STR BETA SMB UMD HML RMW IA	MKTRF SMB BAB	BETA STR IVOL IMD	BETA STR IVOL UMD MKTRF BAB
> 0.9	LTR	LTR						SUE				LTR
> 0.7	BAB SUE					BETA BAB SMB	UMD	UMD		BETA	SMB	
In-Sample α	-0.36	0.36	0.37		-0.07	0.20	0.41	0.03	0.21	0.07	0.91	-0.10
Out-of-Sample α	0.45	0.33	0.14		1.12	0.18	0.38	0.08	0.08	-0.17	-0.13	-0.09

Prob.	N20	N21	N22	N23	N56	N57	N29	N30	N62	N63
	Panel B: I(SVAR $\leq$ 0.4)I(ME $>$ 0.6)				Panel D: I(SVAR $>$ 0.8)					
> 0.95	MKTRF IVOL HML SMB UMD	MKTRF HML SMB STR BETA BAB ROE HMLM	MKTRF SMB UMD	MKTRF SMB HMLM LIQ RMW ROE STR		UMD MKTRF SMB	MKTRF IVOL	MKTRF IVOL STR	MKTRF IVOL SMB	IVOL SMB STR QMJ
> 0.9										
> 0.7			BAB	REG				NI	NI	MKTRF
In-Sample α	0.07	-0.22	0.12	-0.24	1.61	-1.05	0.25	1.20	-0.88	-0.85
Out-of-Sample α	0.12	0.14	-0.02	-0.10	1.68	0.02	0.38	0.38	-0.14	0.27

Table 2 summarizes the factor selection patterns in the mega clusters based on the splits in the top two layers. Panel A (i.e., characteristics-managed asset clusters N8, N18, and N19) contains stocks of low-variance and small-cap (i.e., SVAR ≤ 0.4 and ME ≤ 0.6). We found that MKTRF, IVOL, and BETA are crucial for those stocks, as they are all selected with a probability larger than 0.95. However, the factors selected in N8 differ from those selected in N18 and N19. Figure 5 shows that low stock valuation

(EP, earning-to-price) differentiates N8 from N18 and N19. Momentum (UMD) and short-term reversal (STR) factors are useful in N18 and N19, but do not help price the low-valuation assets in N8.

Panel B includes clusters that contain stocks of low-variance and large-cap (i.e., SVAR ≤ 0.4 and ME > 0.6). MKTRF and SMB are the most important factors for sub-clusters, but each requires its own set of factors. For instance, UMD is selected in N20 and N22 with a probability greater than 0.95, but not in N21 and N23 with a probability greater than 0.7. N21 has clear-cut properties, containing stocks with the largest size and smallest volatility (SVAR < 0.2 and ME > 0.8), and for those stocks, it seems that we need a relatively large number of factors, including the two value factors HML and HMLM.

Panel C consists of 8 clusters with medium variance stocks ($0.4 <$ SVAR ≤ 0.8). These clusters exhibit significant heterogeneity, requiring distinct factors. Panel D includes 6 clusters with the highest volatility stocks (SVAR > 0.8), and involves fewer factors than others. Remarkably, IVOL proves to be an important factor in 4 out of 6 clusters. Because high volatility correlates with high holding costs, larger mispricing (excess alpha) may exist in these clusters (Pontiff, 2006), which is empirically observed. Among these clusters, N56 is unique and contains microcap and past loser stocks characterized by high volatility and low valuation (growth stocks). None of the factor candidates appears to be useful in explaining these stock returns.

Finally, we present in-sample and out-of-sample alphas derived from fitting the cluster-specific factor models. The in-sample alpha represents the average of MCMC posterior samples of the parameter, while the out-of-sample alpha denotes the average of pricing residuals. These monthly alpha values are estimated for individual stock returns and can potentially be relatively high. In Panel B, most alpha values are small for low-variance and large-cap stocks. However, the leaf cluster N56, which does not utilize any of the selected factors, exhibits the highest alphas in-sample (1.61%) and out-of-sample (1.68%). By implementing a long-short alpha strategy that equally weights portfolios in N56 (with the most positive in-sample alpha) and N62 (with the most negative in-sample alpha), hedging corresponding selected factors, we observe

in-sample and out-of-sample return spreads of 2.49% and 1.82%, respectively.

Within-group estimations and tests. Berk (2000) points out the potential errors-in-variables problem in testing factor models using characteristic-sorted portfolios, a common procedure used in the literature (e.g., Litzenberger and Ramaswamy, 1980; Daniel and Titman, 1997). The uncommon factors and different loadings on common factors indicate that the distribution of β across clusters, unlike what Berk (2000) assumes. We therefore demonstrate that examining the variation within groups can help estimation and evaluation, when the variation differs significantly across the sorted groups, and the number of clusters is not large. Note that any errors-in-variables problem only biases against our findings. Within-group tests are valid and useful under appropriate sorting!

Overall, we find strong evidence supporting the existence of asset heterogeneity and uncommon factors. A factor with a negligible selection probability (and thus "weak") in one cluster could be selected with a high probability in another, challenging the standard two-pass procedure for risk premia inference. Instead of using a supervised principal component analysis to estimate the risk premia of weak factors (Giglio et al., 2022) where test assets are chosen based on their correlations with given factors, we jointly select test assets and factors to emphasize asset heterogeneity and resolve the weak identification issue.

4.3 Grouped Heterogeneity in Disguise

Mostly useful and mostly useless factors. We next revisit the tree-growing process and investigate how the importance of factors changes as the tree grows iteratively and generates finer clusters. In each iteration, we define a factor with a selection probability higher than 0.7 in more than 60% of all leaf clusters to be "mostly useful," a factor with a selection probability lower than 0.3 in more than 60% of leaf clusters to be "mostly useless," and the rest to be "partially useful/useless." A factor can move across categories during the iterative tree growth. Table 3 summarizes the results, where ticks indicate mostly useful factors, crosses, mostly useless factors, and circles, partially useful factors.

Table 3: Factor Selection Evaluation Through Iterative Splits

This table shows the iterative factor selection results during tree growth (first column) for an increasing number of clusters. A factor is displayed with a "√" for a selection probability higher than 70% in more than 60% of leaf clusters. A factor is displayed with a "×" for a selection probability lower than 30% in more than 60% leaf clusters. A factor belonging to either group is displayed with an "o".

Split Order	MKTRF	SMB	HML	RMW	CMA	UMD	HMLM	QMJ	BAB	LIQ
Root	√	√	√	√	×	√	√	√	√	√
S1	√	√	×	o	×	√	o	o	√	o
S2	√	√	×	×	×	√	√	×	√	×
S3	√	√	o	×	×	√	√	o	√	×
S4	√	√	√	×	×	√	√	×	√	×
S5	√	√	×	×	×	√	o	×	√	×
S6	√	√	o	o	×	√	o	×	√	×
S7	√	√	o	o	×	√	×	×	√	×
S8	√	√	o	o	×	√	×	×	o	×
S9	√	√	o	o	×	√	×	×	o	×
S10	√	√	o	o	×	√	o	×	o	×
S11	√	√	o	o	×	o	o	×	o	×
S12	√	√	o	o	×	√	×	×	o	×
S13	√	√	o	o	×	o	o	×	o	×
S14	√	√	o	o	×	o	×	×	o	×
S15	√	√	o	o	×	o	×	×	o	×
S16	√	√	o	o	×	o	×	×	o	×
S17	√	√	o	×	×	o	×	×	o	×
S18	√	√	o	×	×	o	×	×	o	×
S19	√	√	o	×	×	o	×	×	o	×
S20	√	√	o	×	×	o	×	×	o	×

Split Order	STR	LTR	REG	IVOL	SUE	IA	ROE	IMD	NI	BETA
Root	√	√	√	√	√	×	√	√	√	√
S1	√	×	√	√	o	o	o	o	√	√
S2	√	×	√	√	×	×	×	×	√	√
S3	√	o	×	√	×	×	o	×	√	√
S4	√	×	×	√	×	×	×	×	√	√
S5	√	×	×	√	×	×	o	×	√	√
S6	o	×	×	√	×	×	o	×	√	o
S7	√	×	×	√	×	×	o	×	o	o
S8	√	×	×	√	×	×	o	×	o	o
S9	√	×	×	√	×	×	×	×	×	o
S10	√	×	×	o	×	×	o	×	o	o
S11	√	×	×	o	×	×	o	×	o	o
S12	√	×	×	o	×	×	o	×	o	o
S13	√	×	×	o	×	×	o	×	o	o
S14	√	×	×	√	×	×	o	×	×	o
S15	√	×	×	o	×	×	o	×	×	o
S16	√	×	×	o	×	×	o	×	×	o
S17	√	×	×	o	×	×	×	×	×	o
S18	√	×	×	o	×	×	×	×	×	o
S19	√	×	×	o	×	×	×	×	×	o
S20	√	×	×	o	×	×	×	×	×	o

At the root node (whole universe of stock returns), 18 out of 20 factors are selected as the most useful. However, the marginal likelihood is the smallest at the root node, as seen in Figure A.1. The marginal likelihood increases as the tree grows, and some factors become mostly useless. For example, at the third split, with four leaves, RMW, CMA, LIQ, REG, SUE, IA, and IMD become mostly useless. When the tree stops growing after 20 splits, MKTRF, STR, and SMB consistently constitute useful factors.

UMD and IVOL are mostly useful in the first half of splits, while CMA and IA are never mostly useful. Furthermore, RMW, QMJ, LIQ, LTR, SUE, and IMD are mostly useful only at the root. Those results are consistent with what we have observed in Tables 1 and 2. Although some factors are mostly useless, they can still be important for some specific clusters. For example, two investment factors (CMA and IA) are mostly useless and have a cross almost at every split, but IA is still helpful in pricing assets in cluster N52, as shown in Table 2.

Overall, BCM highlights how nonlinear interactions in characteristics help identify grouped heterogeneity and how such heterogeneity, when not considered, manifests itself as anomalies previously documented. Incorporating both high-dimensional, nonlinear information from characteristics and the structure of linear factor models, BCM adds to studies pitching firm characteristics versus factors (e.g. Daniel and Titman, 1997) and mitigates the "factor zoo" problem.

Anomalies and associated factors. Grouped heterogeneity may also be a key driver for the anomalies documented in the literature. For example, numerous anomalies have been associated with return volatility and size (e.g., Banz, 1981; Ali, Hwang, and Trombley, 2003; Ang, Hodrick, Xing, and Zhang, 2006). Our framework has identified the prominence of SVAR and ME as asset characteristics for heterogeneity across endogenously generated groups. This result suggests that these anomalies may reflect heterogeneous factor structures across assets, despite small stocks usually having high return variance. Mechanically, it is also possible that a better estimation of the variance instead of the mean improves the marginal likelihood of splitting the cross section.

However, adding a common factor constructed from a particular anomaly may not eliminate the anomaly, particularly if the factor is weak. For example, most low-variance and large-cap stocks do not load on the idiosyncratic volatility (IVOL) factor. In contrast, most high-return-variance stocks do not load on Betting-Against-Beta (BAB) factor. Moreover, a seeming anomaly could result from factors not related to it. For example, the factors the large-cap low-variance clusters (N20, N21, N22, N23) load on greatly differ from those of high-return-variance clusters (N56, N57, N29, N30, N62, N63), and most factors are unrelated to beta or idiosyncratic volatility anomalies.

This observation aligns with Chaieb, Langlois, and Scaillet (2021), which suggests that characteristics are not necessarily paired more often with their corresponding factors.

4.4 Sources for Time Variation in Factor Loadings

Conditional factor models enable us to examine the drivers for the time-varying factor loadings across leaf clusters. Table 4 summarizes the signs and significance of the conditioning variables used for the time-varying factor loadings of MKTRF and SMB, the two most important factors across all clusters, as seen in Table 2. Table 4 has four panels showing the top two layers of the tree. "+" indicates significant and positive predictors, "-", significant and negative predictors, and "○", insignificant predictor. The uncertainty of parameter estimates is based on 2,000 MCMC posterior samples, with 95% credible intervals.

Concerning the loading on MKTRF, we find that market equity (ME) has a persistent positive impact on most leaf clusters' time-varying beta on the market factor, in line with economic intuition: a stock with a higher market equity value has greater exposure to the value-weighted aggregate market. We also learn that the book-to-market (BM) ratio is an important determinant of the time variation of beta on the market factor. Furthermore, investment growth (AGR) is a positive indicator in Panel A, B, and C; however, it has little impact on high return variance stocks in Panel D. In addition, momentum (MOM) is a positive indicator for medium and high return variance stocks in Panel C and D; in contrast, it seems to be a negative indicator for low-variance and small-cap stocks in Panel A, consistent with the findings of Kelly, Moskowitz, and Pruitt (2021) on past performance predicting future market beta.

Concerning loadings on SMB, our analysis uncovers that market equity negatively affects SMB beta in Panels B, C, and D. However, it can positively influence low-variance small-cap stocks in Panel A, aligning with economic intuition. BCM tree reveals a nonlinear relationship between SMB factor loading and firm market equity, which may interact with return variance. Another significant characteristic influencing the time variation of SMB beta is operating profitability (OP). OP positively affects stocks with low variance and large market capitalization and stocks with medium variance in Panels B and C.

Table 4: Significant Characteristics of Factor Loadings

This table explains the conditioning variables' signs and significance for the time-varying factor loadings of the market and beta factors. A "+" indicates a positive significant predictor, a "-" indicates a negative significant predictor, and an "o" indicates an insignificant predictor. The significance is determined from 2,000 MCMC samples' 95% credible parameter intervals. Four panels display leaf clusters' positions under the top three layers of the tree.

	MKTRF Beta Characteristics					SMB Beta Characteristics				
Leaf Node	ME	BM	OP	AGR	MOM	ME	BM	OP	AGR	MOM
				Panel A: $\mathbb{1}\{\text{SVAR} \leq 0.4\}\mathbb{1}\{\text{ME} \leq 0.6\}$						
N8	+	o	o	o	o	o	o	o	o	o
N18	+	+	+	+	-	o	o	o	o	o
N19	+	o	o	o	o	+	o	o	o	-
				Panel B: $\mathbb{1}\{\text{SVAR} \leq 0.4\}\mathbb{1}\{\text{ME} > 0.6\}$						
N20	+	+	+	+	o	-	o	+	+	o
N21	+	+	+	o	o	-	o	+	o	o
N22	o	o	o	+	o	-	o	o	o	o
N23	+	+	+	o	o	-	+	+	o	+
				Panel C: $\mathbb{1}\{\text{SVAR} > 0.4\}\mathbb{1}\{\text{SVAR} \leq 0.8\}$						
N48	o	o	o	o	o	o	o	o	o	o
N49	+	o	o	o	o	o	o	o	o	o
N50	+	+	o	o	o	o	o	o	o	o
N51	+	o	o	o	o	o	o	+	o	o
N52	+	+	+	+	+	-	o	+	+	o
N53	+	o	o	+	o	o	o	o	+	o
N54	o	o	o	o	o	o	o	o	o	o
N55	+	o	o	o	o	o	o	o	o	o
				Panel D: $\mathbb{1}\{\text{SVAR} > 0.4\}\mathbb{1}\{\text{SVAR} > 0.8\}$						
N56	o	o	o	o	o	o	o	o	o	o
N57	+	o	o	o	o	-	o	o	-	o
N29	+	o	o	o	+	o	o	o	o	o
N30	+	+	+	o	o	o	o	o	o	o
N62	+	+	+	o	+	-	o	o	o	+
N63	o	o	o	o	o	o	o	o	-	+

4.5 Pricing Performance and Investment Strategies

An asset pricing model has to be evaluated in terms of its empirical performance. In the standard approach, missing a useful factor or including a useless one may result in a significant alpha. In contrast, BCM offers the advantage of fitting individual stocks in different clusters separately, allowing for heterogeneous factor selection and estimation. Because BCM relies on economically guided statistical fitness (marginal likelihood of factor models), its economic fitness in terms of minimized pricing errors

is not guaranteed and depends on the validity of the assumed grouped heterogeneity with uncommon factors.

To evaluate the pricing error in a specific cluster (j), we rely on the Cross-Sectional R^2 by aggregating individual stocks in the same cluster, defined as

$$\text{Cross-Sectional } R^2 = 1 - \frac{\sum_{i=1}^{N_j}\left(\frac{1}{T_i}\sum_{t=1}^{T_i}(r_{i,t}^{(j)} - \widehat{r}_{i,t}^{(j)})\right)^2}{\sum_{i=1}^{N_j}\left(\frac{1}{T_i}\sum_{t=1}^{T_i} r_{i,t}^{(j)}\right)^2}, \tag{13}$$

where $\widehat{r}_{i,t}^{(j)} = \sum_{j=1}^{J} \mathbb{1}_{\{\widehat{T}(\mathbf{z}_{i,t-1},\mathbf{x}_{t-1})=j\}}\widehat{\boldsymbol{\beta}}_j(\mathbf{z}_{i,t-1})^\intercal \mathbf{f}_t^{(j)}$, and N_j is number of assets in the j-th cluster. It measures how well the model explains cross-sectional differences in average returns across all individual stocks in cluster j. We compare the performance of the heterogeneous factor model with the FF5 model in each of the 21 clusters. A higher Cross-Sectional R^2 value suggests lower pricing errors, and a positive value confirms that the factor model has better fitness than the constant model. We use the FF5 model and its variants as our benchmark because it is arguably the most powerful and widely adopted. The comparison results with other benchmarks are similar.

Table 5 shows that the uncommon-factor models outperform the cluster-specific and pooled FF5 models in both in-sample and out-of-sample analyses. The effectiveness of FF5 models can vary depending on whether cluster-specific or pooled approaches are used. The former takes advantage of cross-sectional cluster homogeneity, while the latter benefits from a significantly larger sample size. Our Skeptic or Agnostic model consistently generates higher Cross-Sectional R^2 values than the FF5 model for almost all clusters. Small-variance and large-cap stocks in Panel B exhibit the lowest pricing errors, whereas high return variance stocks in Panel D display high pricing errors. N56, consisting of microcap, past loser stocks with high volatility and low valuation, remains unexplained by any factor, including those in Table 2. The FF5 model performs even worse for this cluster, suggesting the inclusion of irrelevant factors. Recall that for agnostic investors choose and $w_i = 0.1$ for skeptic investors. Comparing the performance of skeptic and agnostic beliefs, we observe that the latter outperforms in several clusters by a slight margin.

The BCM improvement in asset pricing performance comes from both Bayesian

factor selection and characteristic-based clustering. The latter captures grouped heterogeneity, and the former reflects the asset-pricing implications of the grouped heterogeneity. Identifying the heterogeneity group is extremely useful because BCM informs us about the explanatory power of different assets. Some clusters can be understood by a model with selected factors, while others are very noisy. Note that the Cross-Sectional R^2 for the "well-behaved" clusters can be as high as 25%-27% out-of-sample, which more than doubles that in most other factor models. For example, FF5 without tree clustering, as one of the best-performing factor models in the literature, achieves an R^2 of 11.1% in the full test sample (2007-2021), whereas an FF5 model with BCM clustering (but without factor selection) has improvements for almost all clusters and can achieve an R^2 over 20% out-of-sample for several clusters.

Investment performance. The outperformance of BCM relative to benchmark models such as FF5 in fitting each cluster with smaller pricing errors suggests that uncommon-factor models can provide more accurate expected returns and covariance estimates. Using these model-implied parameters to form optimal portfolios can improve investment performance, further validating and evaluating BCM. Factor models are also routinely used for developing trading strategies in both active and passive investing and for risk management. Therefore, it is natural to examine how BCM helps with investments.

To evaluate the performance of investment strategies using portfolios created for each cluster based on the selected factors, we examine the value-weighted portfolios of the 21 leaf clusters depicted in Figure 5, and the Fama-French 5×5 Size-BE/ME portfolios. Table 6 presents the performance of two portfolio strategies: the LS-Cluster strategy, which longs the leaf cluster portfolio with the highest factor model implied returns and shorts the lowest one, and the MVE-Cluster strategy, which is the mean-variance efficient portfolio on factor model implied returns of all clusters. Both BCM approaches (Skeptic and Agnostic) outperform the FF5 model in and out of the sample in the LS-Cluster strategy, i.e., LS-Cluster strategies based on BCM generate higher Sharpe ratios and CAPM alphas than that on the FF5. Furthermore, the MVE-Cluster strategy based on BCM with agnostic priors outperforms in- and out-of-sample, pro-

viding an annualized Sharpe ratio of 1.8 and a monthly CAPM alpha of 0.86%.

Notice that regardless of the factor models we use, the out-of-sample Sharpe ratio of investment portfolios constructed from the BCM clusters (e.g., 1.8 and 1.64 with MVE-Cluster using Agnostic and FF5, respectively) can easily triple that of portfolios constructed using the same strategy but from FF25 portfolios (0.6 and 0.52). The out-performance in CAPM Alpha is equally strong, both in- and out-of-sample. Relative to other benchmark investment portfolios constructed from single-sorted portfolios, or common risk-factor premia, the improvements from BCM clustering are even bigger.

Table 5: Evaluating Pricing Performance with Leaf Clusters

This table presents the cross-sectional R^2 (%) values for different models, computed for each leaf cluster to evaluate pricing errors in fitted factor models. Skeptic ($w_i = 0.1$) and agnostic ($w_i = 0.5$) prior models for BCM are shown. We include Fama-French five factors in both cluster-specific and pooled samples. The leaf node IDs match the tree cluster in Figure 5.

	In-Sample (1972-2006)				Out-of-Sample (2007-2021)			
Leaf Node	Skeptic	Agnostic	FF5-Cl	FF5-Pl	Skeptic	Agnostic	FF5-Cl	FF5-Pl
				Panel A: I(SVAR $\leq$ 0.4)I(ME $\leq$ 0.6)				
N8	12.2	13.1	11.4	13.6	14.5	18.3	12.6	16.5
N18	13.0	13.3	11.8	14.0	14.4	19.3	16.1	21.0
N19	14.5	14.6	12.7	12.3	16.6	17.4	16.2	18.0
				Panel B: I(SVAR $\leq$ 0.4)I(ME $>$ 0.6)				
N20	21.0	21.0	19.5	18.9	18.0	17.7	15.5	15.8
N21	15.0	14.9	13.7	14.9	25.1	26.8	24.4	26.5
N22	16.1	16.9	15.9	17.1	19.4	19.4	17.6	19.0
N23	15.2	15.8	14.8	15.9	15.6	16.7	12.6	14.5
				Panel C: I(SVAR $>$ 0.4)I(SVAR $\leq$ 0.8)				
N48	11.7	12.9	11.2	11.8	10.3	15.9	11.2	12.9
N49	17.6	17.7	16.0	17.0	17.4	19.4	18.3	18.4
N50	13.2	14.1	13.0	13.0	11.8	13.8	10.7	11.0
N51	12.9	13.4	12.3	12.7	14.8	16.7	13.1	13.0
N52	15.4	15.4	14.3	14.0	19.4	21.4	18.6	18.4
N53	15.3	15.5	14.4	14.1	16.9	17.3	15.0	15.2
N54	19.0	20.2	17.4	16.7	23.9	25.0	22.6	22.1
N55	11.8	11.7	9.4	10.3	13.6	14.3	11.5	12.2
				Panel D: I(SVAR $>$ 0.8)				
N56	14.2	17.5	12.9	11.4	4.7	5.5	4.0	3.7
N57	16.5	16.3	11.5	11.3	14.6	16.6	13.2	11.9
N29	12.8	12.7	11.2	10.9	11.3	11.4	9.5	9.2
N30	17.9	18.9	16.8	17.2	22.8	24.4	19.4	18.9
N62	12.9	13.5	10.6	11.0	14.1	15.5	12.6	13.4
N63	15.6	16.0	14.0	13.9	11.9	13.2	10.2	11.1

Table 6: Investing in Cluster-Selected Factor Models

This table presents the investment performance of value-weighted portfolios corresponding to the 21 leaf clusters depicted in Figure 5. Skeptic ($w_i = 0.1$) and agnostic ($w_i = 0.5$) indicate different prior beliefs of factor usefulness by investors. The "LS-Cluster" longs the leaf cluster portfolio with the highest factor model implied returns (loadings on risk premia) and shorts the lowest. The "MVE-Cluster" strategy is the mean-variance efficient portfolio on cluster-selected factor model implied returns. We have also provided results for the Fama-French ME-B/M 5×5 portfolios.

	In-Sample (1972-2006)			Out-of-Sample (2007-2021)		
Portfolio	Skeptic	Agnostic	FF5	Skeptic	Agnostic	FF5
			Panel A: Annualized Sharpe Ratio			
LS-Cluster	1.14	0.94	0.55	0.45	0.24	0.20
MVE-Cluster	0.81	0.88	0.82	1.23	1.80	1.64
LS-Cluster (ME-B/M 5×5)	0.90	0.89	0.79	0.19	0.27	-0.11
MVE-Cluster (ME-B/M 5×5)	0.70	0.83	0.80	0.53	0.60	0.52
			Panel B: CAPM Alpha (%)			
LS-Cluster	2.30***	2.04***	1.16***	1.18	0.75	0.69
MVE-Cluster	0.50***	0.58***	0.46***	0.79***	0.86***	0.75***
LS-Cluster (ME-B/M 5×5)	1.21***	1.43***	0.89***	0.06	0.04	-0.64*
MVE-Cluster (ME-B/M 5×5)	0.53***	0.60***	0.69***	-0.16	-0.06	-0.21

5 Macro-Instrumented Asset Pricing Regimes

The macroeconomic environment continuously evolves, and investors learn from experience and data to adjust their strategies. The underlying return dynamics and, thus, the factor structure are unlikely to stay the same. The finance literature has documented extensive evidence of model instability and breaks. For example, Smith and Timmermann (2021, 2022) examine the risk of structural breaks and find abrupt changes in factor loadings. Without factor selection or time-varying loadings, structural breaks not only lead to inconsistent estimates and model misspecifications, but also inflate the number of factors (Pelger and Xiong, 2022). As described in Section 2, BCM can account for structural breaks via clustering time series observations. BCM identifies macroeconomic variables (e.g., market volatility, inflation, or treasury bill rate) that reflect time-series asset heterogeneity, providing insights into the factors driving returns under various macro-instrumented asset-pricing regimes.

Time-series splits. A time-series split works similarly to a cross-sectional split described in Section 4. The algorithm searches for split candidates in time-series (macro) and cross-sectional (characteristics) variables. Either type of split candidate partitions the panel of asset returns into two child leaves. The tree-growing process proceeds similarly, with the marginal likelihood in (5) being evaluated iteratively. By jointly comparing time-series and/or cross-sectional candidates at each split, BCM determines the optimal type of split and identifies both cross-sectional and time-series heterogeneity of asset returns.

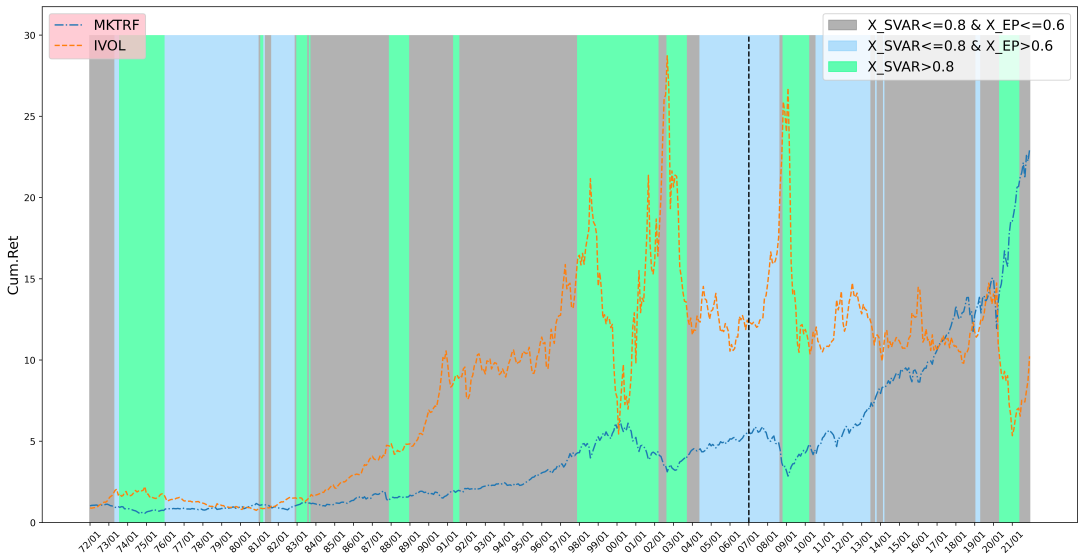
In our empirical implementation, we impose an economic restriction that the algorithm splits the sample first using macroeconomic variables, with up to two splits that result in at most three regimes for illustration and easy interpretation, and then using asset-specific characteristics. Note that the regimes are defined according to the value of macro variables, i.e., macro-instrumented. Thus, they generally correspond to discontinuous episodes in calendar time. As shown in Figure 6, three regimes indicated by two macroeconomic states, market volatility (X_SVAR) and market earnings-to-price (X_EP), emerge from the training sample from January 1972 to December 2006: (i) High-Vol with X_SVAR > 0.8, (ii) Non-High Vol and High Valuation with X_SVAR $\leq$ 0.8 and X_EP > 0.6, and (iii) Non-High-Vol and Low-Valuation with X_SVAR ≤ 0.8 and X_EP ≤ 0.6. Based on the in-sample estimates, we analyze the regimes in the test sample (also displayed in the figure). The cumulative returns of the two most frequently selected factors, IVOL and MKTRF, are superimposed and can be related to the market valuation and volatility dynamics.

Figure A.2 (in the appendix) presents the corresponding BCM tree and asset clusters, with macro predictors selected by their ranks scaled to $[0, 1]$ based on the previous 10-year rolling window. Market volatility is the most critical predictor of regime switching, followed by market EP in non-high-volatility periods, all consistent with Pastor and Stambaugh (2001), which features breaks in a model linking equity risk premia to changes in stock market volatility.

Regime-specific factor models. Following the time series splits, three subtrees split the cross section by firm characteristics under those three macro-instrumented as-

Figure 6: Time-Series Breaks

Macro-instrumented regimes in calendar time (tree plotted in Figure A.2 in the appendix), with multiple regime switches corresponding to the first two splits based on market volatility and earnings-to-price. The training data is from 1972-2006, and the test data is from 2007-2021. The cumulative returns of the two most commonly selected factors are included.



set pricing regimes. Individual stock return variance remains the first characteristic that splits the cross section under each regime, forming asset clusters under the high-volatility regime based on individual stock return variance. Remarkably, the firm characteristics and thresholds used for cross-sectional splits are the same in the top two layers under the low- and high-valuation regimes. However, the resulting asset clusters have very different factor structures. In total, 21 asset clusters are discovered. The tree that accounts for cross-sectional and time-series asset heterogeneity differs greatly from that of Figure 5, which leaves out time-series breaks.

Table 7 presents the selection of uncommon factors for each asset cluster, considering both cross-sectional and time-series asset heterogeneity. The table includes three panels corresponding to the three macro-instrumented regimes. Asset clusters in the high-volatility regimes (Panel C) have fewer selected factors than others (Panels A and B), with IVOL being the most crucial factor. In Panel A, which represents the non-high-volatility and low-valuation regime, MKTRF is frequently selected along with IVOL. In the high-valuation regime (non-high volatility and high valuation) of Panel C, profitability and investment factors also appear important. Furthermore, the

high-volatility regime exhibits the largest discrepancy between the out-of-sample and in-sample alphas, indicating the poor performance of the underlying factor models during periods of high market volatility. Cluster N30 is characterized by high return variance and low cash flow during periods of high market volatility and shows a substantial difference between out-of-sample and in-sample alphas.

We then examine the asset pricing performance of cluster-specific factor models under each regime. Table 8 presents the results. Except for the cross-sectional R^2 defined in (13), we also report the total R^2, which captures the fraction of time-series variations of realized equity returns explained by the contemporaneous factors, aggregated over time and assets, defined as:

Table 7: Regime Switches: Factor Selection Evaluation by Leaf Clusters

This table presents the uncommon factor selection outcomes for the cross-sectional clustered model during regime changes (tree plotted in Figure A.2 in the appendix). The rows labeled > 0.95, > 0.9, and > 0.7 correspond to different selection probability thresholds, and the listed factors are ordered based on their values. Three panels display leaf cluster positions for the three macroeconomic-targeted regimes. We present in-sample (1972-2006) and out-of-sample (2007-2021) alphas (in %) for cluster-specific fitted factor models.

Panel A: I($X_SVAR \leq 0.8$) I($X_EP \leq 0.6$)									
Prob.	N32	N33	N34	N35	N36	N74	N75	N38	N39
>0.95	STR IVOL	MKTRF IVOL IMD SUE	MKTRF SMB STR IA	IMD	QMJ IVOL SUE	HML IVOL MKTRF CMA	IVOL IMD BETA	IVOL	IVOL MKTRF
>0.9					HML	SUE		LTR	
>0.7	HML QMJ BAB	NI	IVOL LTR						NI
In-Sample α	0.09	1.05	0.02	0.18	1.70	0.59	0.55	0.70	0.03
Out-of-Sample α	-0.05	0.60	0.06	0.43	0.50	-0.14	-0.08	0.26	0.70

	Panel B: I($X_SVAR \leq 0.8$) I($X_EP > 0.6$)						Panel C : I($X_SVAR > 0.8$)					
Prob.	N20	N42	N43	N44	N45	N23	N12	N13	N28	N29	N30	N31
>0.95	STR IVOL SUE ROE	MKTRF IVOL IA BETA	IVOL SUE ROE IMD	MKTRF IVOL BETA	QMJ LIQ STR IVOL SUE	SMB RMW QMJ IVOL SUE IA IMD BETA	MKTRF IVOL SUE	BETA	HMLM STR IVOL	STR IVOL IA	IVOL BETA	STR IVOL
>0.9									LTR			
>0.7		ROE		UMD	NI			IVOL				
In-Sample α	1.12	0.36	0.36	0.82	1.41	-0.36	0.41	0.66	0.66	0.69	0.89	0.17
Out-of-Sample α	0.10	1.62	0.44	1.42	1.09	0.05	0.64	1.53	2.25	2.22	6.02	2.06

Table 8: Asset Pricing Performance Across Regimes

This table compares the asset pricing performance of various models based on three metrics in different macroeconomic regimes. It displays models with skeptical and agnostic priors for BCM. We include the Fama-French five-factor model in cluster-specific and pooled samples. The clustering model is based on the three clusters (tree plotted in Figure A.2 in the appendix).

	In-Sample (1972-2006)				Out-of-Sample (2007-2021)			
				Panel A: I(X_SVAR $\leq$ 0.8) I(X_EP $\leq$ 0.6)				
	Skeptic	Agnostic	FF5-Cl	FF5-Pl	Skeptic	Agnostic	FF5-Cl	FF5-Pl
Total R^2 (%)	13.3	13.3	12.2	11.3	15.4	15.3	14.1	13.2
Cross-Sectional R^2 (%)	11.0	11.3	9.6	9.4	15.1	16.3	14.3	13.5
Predictive R^2 (%)	0.4	0.5	0.2	0.1	-0.1	-0.1	-0.2	-0.2
				Panel B: I(X_SVAR $\leq$ 0.8) I(X_EP $>$ 0.6)				
	Skeptic	Agnostic	FF5-Cl	FF5-Pl	Skeptic	Agnostic	FF5-Cl	FF5-Pl
Total R^2 (%)	22.5	22.3	21.4	19.8	11.8	12.0	10.9	9.5
Cross-Sectional R^2 (%)	34.0	33.9	31.4	30.6	11.6	10.7	9.7	7.6
Predictive R^2 (%)	1.1	1.1	1.1	1.0	0.2	0.2	0.1	0.1
				Panel C: I(X_SVAR $>$ 0.8)				
	Skeptic	Agnostic	FF5-Cl	FF5-Pl	Skeptic	Agnostic	FF5-Cl	FF5-Pl
Total R^2 (%)	21.3	21.0	17.7	15.3	17.5	15.6	12.5	11.2
Cross-Sectional R^2 (%)	11.7	21.2	19.6	16.9	18.0	33.4	29.0	28.4
Predictive R^2 (%)	0.6	0.2	0.3	0.3	0.2	0.6	0.6	0.8

$$\text{Total } R^2 = 1 - \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(r_{i,t} - \widehat{r}_{i,t})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T_i} r_{i,t}^2}, \tag{14}$$

where $\widehat{r}_{i,t} = \sum_{j=1}^{J} \mathbb{1}_{\{\widehat{T}(\mathbf{z}_{i,t-1},\mathbf{x}_{t-1})=j\}} \widehat{\boldsymbol{\beta}}_j(\mathbf{z}_{i,t-1})^\intercal \mathbf{f}_t$, and the predictive R^2, defined as:

$$\text{Predictive } R^2 = 1 - \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i}(r_{i,t} - \widehat{r}_{i,t})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T_i} r_{i,t}^2}, \tag{15}$$

where $\widehat{r}_{i,t} = \sum_{j=1}^{J} \mathbb{1}_{\{\widehat{T}(\mathbf{z}_{i,t-1},\mathbf{x}_{t-1})=j\}} \widehat{\boldsymbol{\beta}}_j(\mathbf{z}_{i,t-1})^\intercal \boldsymbol{\lambda}_\mathbf{f}$, and $\boldsymbol{\lambda}_\mathbf{f}$ is the factor risk premia estimate, a simple average return of the tradable factor. Predictive R^2 captures the predictive performance of model-implied return forecasts and summarizes the model's ability to capture compensation for systematic risk exposures.

Unlike in Table 5, the cluster-specific FF5 model, which only uses data in the regime, consistently outperforms the pooled version, reflecting the superior information when considering macro-instrumented regimes (even without factor selection). In

Table 8, the three performance measures suggest that the cluster-specific factor models from our BCM almost always outperform both FF5 models in- and out-of-sample for skeptical and agnostic investors in the low- and high-value regimes, and for agnostic investors in the high-volatility regime. The high-volatility regime is also where factor models explain the most variation in returns, and heterogeneous factor models show the greatest improvement compared to other benchmark (common-)factor models.

6 Conclusion

Sparse modeling aims to provide interpretability and tractability, and the literature thus far assumes a *globally* sparse structure or a small set of *common* covariates applying to all observations. However, potential grouped heterogeneity challenges this approach, partially causing the "illusion of sparsity" (Giannone et al., 2021). We provide an alternative sparse modeling framework for exploiting large panel data and accommodating grouped heterogeneity through joint observation clustering and heterogeneous variable selection. Our BCM tree utilizes a natural global split criterion—maximizing the joint marginal likelihood analytically derived using spike-and-slab priors. As such, BCM prevents overfitting seen in conventional tree-based models, achieves sparsity and economic interpretability, and maintains the advantages of dense modeling (e.g., mitigating model misspecifications, see Avramov et al., 2023).

When estimating uncommon-factor models for cross-sectional asset pricing, BCM combines information from firm characteristics, macroeconomic states, and systematic factors to sparsely model cross-sectional heterogeneity captured by high-dimensional firm characteristics and detect macroeconomic-instrumented asset pricing regimes. We find compelling grouped heterogeneity in the cross section (linked to asset return variance, market cap, and earnings-to-price valuation ratio) and time series (predicted by market volatility and valuation) in U.S. equities. MKTRF and SMB are robust common factors, while a small set of uncommon factors are identified as drivers of returns across characteristics-managed-market-timed clusters. BCM also helps explain volatility- or size-related anomalies, enables testing models within sorted portfolios, and mitigates the "factor zoo" problem, in addition to outperforming benchmark

common-factor models in pricing and investments.

Future studies could explore the theoretical underpinnings of the grouped heterogeneity and uncommon factors found in the data, or apply BCM to other asset classes or across asset classes and countries. Given how BCM embodies a general economically guided, goal-oriented (greedy) search in the large modeling space for clustering panel observations, its applications to panel data with grouped heterogeneity beyond asset pricing and the objective of maximum marginal likelihood can also be promising for general social science research.

References

Ahn, D.-H., J. Conrad, and R. F. Dittmar (2009). Basis assets. *Review of Financial Studies 22*(12), 5133–5174.

Ali, A., L.-S. Hwang, and M. A. Trombley (2003). Arbitrage risk and the book-to-market anomaly. *Journal of Financial Economics 69*(2), 355–373.

Ang, A. and G. Bekaert (2002). International asset allocation with regime shifts. *Review of Financial Studies 15*(4), 1137–1187.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). The cross-section of volatility and expected returns. *Journal of Finance 61*(1), 259–299.

Asness, C. and A. Frazzini (2013). The devil in HML's details. *Journal of Portfolio Management 39*(4), 49–68.

Asness, C. S., A. Frazzini, and L. H. Pedersen (2019). Quality minus junk. *Review of Accounting Studies 24*(1), 34–112.

Athey, S. and G. W. Imbens (2019). Machine learning methods that economists should know about. *Annual Review of Economics 11*, 685–725.

Avramov, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics 64*(3), 423–458.

Avramov, D., S. Cheng, L. Metzker, and S. Voigt (2023). Integrating factor models. *Journal of Finance 78*(3), 1593–1646.

Avramov, D. and T. Chordia (2006). Asset pricing models and financial market anomalies. *Review of Financial Studies 19*(3), 1001–1040.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial economics 9*(1), 3–18.

Barillas, F. and J. Shanken (2018). Comparing asset pricing models. *Journal of Finance 73*(2), 715–754.

Bekaert, G., R. J. Hodrick, and X. Zhang (2009). International stock return comovements. *Journal of Finance 64*(6), 2591–2626.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica 80*(6), 2369–2429.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies 81*(2), 608–650.

Berk, J. B. (2000). Sorting Out Sorts. *Journal of Finance 55*, 407–427.

Bianchi, D., M. Büchner, and A. Tamoni (2021). Bond risk premiums with machine learning. *Review of Financial Studies 34*(2), 1046–1089.

Black, F. (1974). International capital market equilibrium with investment barriers. *Journal of Financial Economics 1*(4), 337–352.

Bonhomme, S. and E. Manresa (2015). Grouped patterns of heterogeneity in panel data. *Econometrica 83*(3), 1147–1184.

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

Bryzgalova, S., M. Pelger, and J. Zhu (2022). Forest through the trees: Building cross-sections of stock returns. Technical report, London Business School.

Bugni, F., I. Canay, A. Shaikh, and M. Tabord-Meehan (2022). Inference for cluster randomized experiments with non-ignorable cluster sizes. *arXiv preprint arXiv:2204.08356*.

Canay, I. A., A. Santos, and A. M. Shaikh (2021). The wild bootstrap with a "small" number of "large" clusters. *Review of Economics and Statistics 103*(2), 346–363.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance 52*(1), 57–82.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010). The horseshoe estimator for sparse signals. *Biometrika 97*(2), 465–480.

Chaieb, I., H. Langlois, and O. Scaillet (2021). Factors and risk premia in individual international stock returns. *Journal of Financial Economics 141*(2), 669–692.

Chib, S. and X. Zeng (2020). Which factors are risk factors in asset pricing? A model scan framework. *Journal of Business & Economic Statistics 38*(4), 771–783.

Chib, S., L. Zhao, and G. Zhou (2023). Winners from winners: A tale of risk factors. *Management Science, Forthcoming*.

Chinco, A., A. D. Clark-Joseph, and M. Ye (2019). Sparse signals in the cross-section of returns. *Journal of Finance 74*(1), 449–492.

Chipman, H. A., E. I. George, and R. E. McCulloch (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics 4*, 266–298.

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

Cong, L. W., G. Feng, J. He, and X. He (2022). Asset Pricing with Panel Tree under Global Split Criteria. Technical report, City University of Hong Kong.

Cong, L. W., K. Tang, J. Wang, and Y. Zhang (2020). AlphaPortfolio: Direct construction through deep reinforcement learning and interpretable AI. Technical report, Cornell Univeristy.

Cui, L., G. Feng, and Y. Hong (2023). Regularized GMM for time-varying models with application to asset pricing. *International Economic Review, Forthcoming*.

Daniel, K. and S. Titman (1997). Evidence on the characteristics of cross sectional variation in stock returns. *Journal of Finance 52*(1), 1–33.

De Bondt, W. F. and R. Thaler (1985). Does the stock market overreact? *Journal of Finance 40*(3), 793–805.

Errunza, V. and E. Losq (1985). International asset pricing under mild segmentation: Theory and test. *Journal of Finance 40*(1), 105–124.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F. and K. R. French (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance 51*(1), 55–84.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–636.

Fan, J., Z. T. Ke, Y. Liao, and A. Neuhierl (2022). Structural deep learning in conditional asset pricing. *Available at SSRN 4117882*.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association 96*(456), 1348–1360.

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

Foerster, S. R. and G. A. Karolyi (1999). The effects of market segmentation and investor recognition on asset prices: Evidence from foreign stocks listing in the United States. *Journal of Finance 54*(3), 981–1013.

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

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

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

George, E. I. and R. E. McCulloch (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association 88*(423), 881–889.

Giannone, D., M. Lenza, and G. E. Primiceri (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica 89*(5), 2409–2437.

Giglio, S., D. Xiu, and D. Zhang (2022). Test Assets and Weak Factors. Technical report, Yale University.

Gospodinov, N., R. Kan, and C. Robotti (2014). Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors. *Review of Financial Studies 27*(7), 2139–2170.

Griffin, J. M. (2002). Are the Fama and French factors global or country specific? *Review of Financial Studies 15*(3), 783–803.

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

Harvey, C. R. (2017). Presidential Address: The Scientific Outlook in Financial Economics. *Journal of Finance 72*, 1399–1440.

Harvey, C. R. and Y. Liu (2021). Lucky factors. *Journal of Financial Economics 141*(2), 413–435.

Hastie, T., A. Montanari, S. Rosset, and R. J. Tibshirani (2022). Surprises in high-dimensional ridgeless least squares interpolation. *Annals of statistics 50*(2), 949.

He, J. and P. R. Hahn (2021). Stochastic tree ensembles for regularized nonlinear regression. *Journal of the American Statistical Association*, 1–20.

He, J., S. Yalov, and P. R. Hahn (2019). XBART: Accelerated Bayesian additive regression trees. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 1130–1138.

He, Z., B. Kelly, and A. Manela (2017). Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics 126*(1), 1–35.

Hou, K., G. A. Karolyi, and B.-C. Kho (2011). What factors drive global stock returns? *Review of Financial Studies 24*(8), 2527–2574.

Hou, K., H. Mo, C. Xue, and L. Zhang (2021). An augmented q-factor model with expected growth. *Review of Finance 25*(1), 1–41.

Hou, K., C. Xue, and L. Zhang (2015). Digesting anomalies: An investment approach. *Review of Financial Studies 28*(3), 650–705.

Ibragimov, R. and U. K. Müller (2010). t-statistic based correlation and heterogeneity robust inference. *Journal of Business & Economic Statistics 28*(4), 453–468.

Jarrow, R. and P. Protter (2016). Positive alphas and a generalized multiple-factor asset pricing model. *Mathematics and Financial Economics 10*(1), 29–48.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance 48*(1), 65–91.

Jensen, M. C., F. Black, and M. S. Scholes (1972). The capital asset pricing model: Some empirical tests.

Kelly, B. T., S. Malamud, and K. Zhou (2023). The virtue of complexity in return prediction. *Journal of Finance 79/1*, 459–503.

Kelly, B. T., T. J. Moskowitz, and S. Pruitt (2021). Understanding momentum and reversal. *Journal of Financial Economics 140*(3), 726–743.

Kleibergen, F. (2009). Tests of risk premia in linear factor models. *Journal of Econometrics 149*(2), 149–173.

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

Levy, H. (1978). Equilibrium in an Imperfect Market: A Constraint on the Number of Securities in the Portfolio. *American Economic Review 68*, 643–658.

Litzenberger, R. H. and K. Ramaswamy (1980). Dividends, short selling restrictions, tax-induced investor clienteles and market equilibrium. *Journal of Finance 35*(2), 469–482.

Loughran, T. and J. R. Ritter (1995). The new issues puzzle. *Journal of Finance 50*(1), 23–51.

Merton, R. C. (1987). A Simple Model of Capital Market Equilibrium with Incomplete Information. *Journal of Finance 42*, 483–510.

Mitchell, T. J. and J. J. Beauchamp (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association 83*(404), 1023–1032.

Nagel, S. and K. J. Singleton (2011). Estimation and evaluation of conditional asset pricing models. *Journal of Finance 66*(3), 873–909.

Neyshabur, B., Z. Li, S. Bhojanapalli, Y. LeCun, and N. Srebro (2019). The role of over-parametrization in generalization of neural networks. In *7th International Conference on Learning Representations, ICLR 2019*.

Pastor, L. and R. Stambaugh (2001). The Equity Premium and Structural Breaks. *Journal of Finance 56*, 1207–1239.

Pástor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*(3), 642–685.

Patton, A. J. and B. M. Weller (2022). Risk price variation: The missing half of empirical asset pricing. *Review of Financial Studies 35*(11), 5127–5184.

Pelger, M. and R. Xiong (2022). State-varying factor models of large dimensions. *Journal of Business & Economic Statistics 40*(3), 1315–1333.

Pontiff, J. (2006). Costly arbitrage and the myth of idiosyncratic risk. *Journal of Accounting and Economics 42*(1-2), 35–52.

Rendleman Jr, R. J., C. P. Jones, and H. A. Latane (1982). Empirical anomalies based on unexpected earnings and the importance of risk adjustments. *Journal of Financial Economics 10*(3), 269–287.

Rossi, A. G. and A. Timmermann (2015). Modeling covariance risk in Merton's ICAPM. *Review of Financial Studies 28*(5), 1428–1461.

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

Smith, S. C. and A. Timmermann (2022). Have risk premia vanished? *Journal of Financial Economics 145*(2), 553–576.

Stock, J. H. and M. W. Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association 97*(460), 1167–1179.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology 58*(1), 267–288.

Van Binsbergen, J. H., X. Han, and A. Lopez-Lira (2023). Man versus machine learning: The term structure of earnings expectations and conditional biases. *Review of Financial Studies 36*(6), 2361–2396.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508.

Online Appendices

I. BCM Algorithm

Section 2.3 presents the step-by-step tree growing examples, while this section illustrates the complete growing algorithm in pseudo-codes.

Algorithm Bayesian Clustering Model

1: **procedure** BCM
2: **Input:** data and parameters. **Output:** a tree structure with a bunch of split rules that defining clusters by panel characteristics.
3: **for** j from 1 to num_iter **do** ▷ Loop over number of iterations
4: **if** current depth $\geq d_{\max}$ **then**
5: **return**.
6: **else**
7: Search the tree, find all leaf nodes $\mathcal{A}$
8: **for** each leaf node s in $\mathcal{A}$ **do** ▷ Loop over all current leaf nodes
9: **if** s is not labeled as "cannot split" **then**
10: **for** each split candidate $\tilde{c}_m$ in $\mathcal{C}_s$ **do**
11: Partition data temporally in N according to $\tilde{c}_m$.
12: **if** Left or right child node cannot satisfy minimal leaf size **then**
13: $L(\tilde{c}_m) = -\infty$.
14: **else**
15: Calculate the split criteria $L(\tilde{c}_m)$ in (9).
16: **end if**
17: **end for**
18: **end if**
19: **end for**
20: Find the best leaf node and split rule that maximizes split criteria

$$\tilde{c}_j = \underset{N \in \mathcal{N}, \tilde{c}_m \in \cup_{s=1}^{|\mathcal{N}|} \mathcal{C}_s}{\arg\max} \{L(\tilde{c}_j)\}$$

21: Split the node selected at the j-th split rule of the tree $\tilde{c}_j$. If a null cutpoint is selected, label the corresponding node "cannot split" and return to line 7.
22: **end if**
23: **end for**
24: **for** j from 1 to J **do** ▷ Loop over all clusters returned by the tree
25: Draw posterior samples of parameters by the Gibbs sampler in Section 2.3.
26: **end for**
27: **return**
28: **end procedure**

Note that the number of clusters J is not determined by users, but is learnt from data after the `for` loop ending in line 23.

II. Proof of Proposition 1

Proof. To simplify notations, we re-express the model in (2) as

$$\mathbf{R} = \mathcal{W}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{\beta} = [\alpha_j, \mathbf{b}_{j,0}^\mathsf{T}, \mathbf{b}_{j,1}^\mathsf{T}]$ and $\mathcal{W} = [1, \mathbf{F}, \mathbf{F} \bullet \mathbf{Z}]$.

First, we keep taking conditions with respect to the assignment to spike or slab $\boldsymbol{\gamma}_j$ and integrate out all other parameters. Given $\boldsymbol{\gamma}_j$, the prior on $\boldsymbol{\beta}$ is $N(0, \sigma_j^2 \boldsymbol{\Lambda}_{0|\boldsymbol{\gamma}_j}^{-1})$, where $\boldsymbol{\Lambda}_{0|\boldsymbol{\gamma}_j}$ is a diagonal matrix with s-th element being $\xi_1^{-2}\sigma_j^{-2}$ if $\gamma_{j,s} = 1$, or $\xi_0^{-2}\sigma_j^{-2}$ if $\gamma_{j,s} = 0$. Following the standard result of Bayesian linear regression with conjugate Normal-Inverse-Gamma prior, integrating out all regression coefficients and residual variance yields,

$$p(\mathbf{R} \mid \boldsymbol{\gamma}_j, \mathbf{Z}, \mathbf{F}) = \frac{1}{(2\pi)^{N/2}} \sqrt{\frac{|\boldsymbol{\Lambda}_{0|\boldsymbol{\gamma}_j}|}{|\boldsymbol{\Lambda}_N|}} \frac{v_0^{S_0}}{v_N^{S_N}} \frac{\Gamma(S_N)}{\Gamma(S_0)},$$

where $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$ is the gamma function.

Second, we integrate out the indicator of assignment $\boldsymbol{\gamma}_j$. Let $\{I_m\}_{m=1}^{2^K}$ denote a collection of all length K vectors with each element taking values 0 or 1 indicating all possible values that $\boldsymbol{\gamma}_{j,\mathbf{f}}$ can take. Note the prior of each model is $\pi(\boldsymbol{\gamma}_{j,\mathbf{f}} = I_m) = \prod_{k=1}^K w_k^{\gamma_{j,\mathbf{f},k}}(1-w_k)^{(1-\gamma_{j,\mathbf{f},k})}$ is defined in (4). The Bayes rule yields,

$$p(\mathbf{R} \mid \mathbf{Z}, \mathbf{F}) = \sum_{I_m} \pi(\boldsymbol{\gamma}_{j,\mathbf{f}} = I_m) p(\mathbf{R} \mid \boldsymbol{\gamma}_j, \mathbf{Z}, \mathbf{F}),$$

The selection of interaction terms $\mathbf{f}_t \otimes \mathbf{z}_{i,t-1}$ follows that of the main effect of $\mathbf{f}$. Therefore, $\boldsymbol{\gamma}_j$ is fixed, given $\boldsymbol{\gamma}_{j,\mathbf{f}}$. □

III. Figure: Marginal (Log) Likelihood Improvements

Figure A.1: This bar plot shows the increase in log marginal likelihood (without normalizing constants) compared to the root model with no splits. To avoid overfitting, the tree applies a penalty (8) during growth and stops after 20 splits, resulting in 21 terminal leaves.

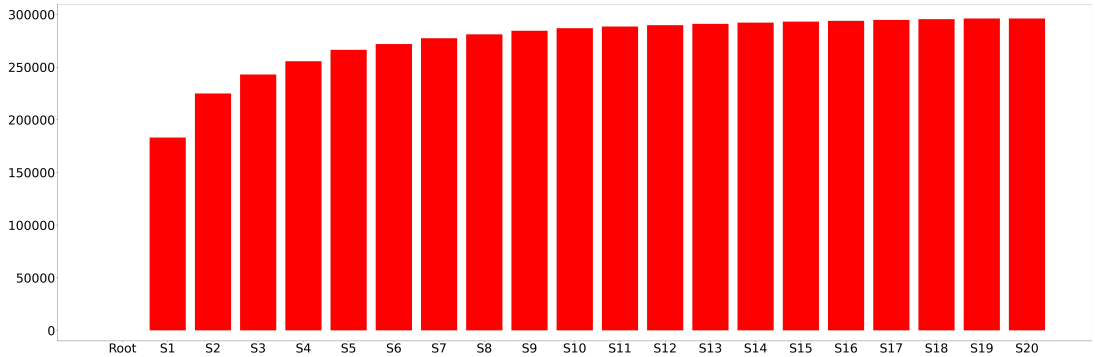


Figure A.2: Tree Clusters Under Macro-Instrumented Regimes

The tree is constructed to partition individual stock returns based on cross-sectional and time-series dimensions, using monthly data from 1972 to 2006. The initial two splits select aggregate predictors, namely market volatility and earnings-to-price, based on their standardized values within a 10-year rolling window, ranging from 0 to 1. Following the time-series splits, the three sub-trees on the cross-section under three regimes are labeled in Figure 5.

