CENTRE FOR ECONOMETRIC ANALYSIS
CEA@BAYES

https://www.bayes.city.ac.uk/faculties-and-research/centres/cea

Bayes Business School (formerly Cass)
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

Time-Varying Factor Selection: A Sparse Fused GMM Approach

Liyuan Cui, Guanhao Feng, Yongmiao Hong, and Jiangshan Yang

CEA@Bayes Working Paper Series

WP–CEA–06-2023

Time-Varying Factor Selection: A Sparse Fused GMM Approach *

Liyuan Cui[†] Guanhao Feng[‡] Yongmiao Hong[§] Jiangshan Yang[¶]

July 31, 2023

Abstract

Empirical asset pricing studies evaluate and select risk factors solely based on their historical aggregate performance, implicitly assuming a time-invariant model specification, and overlooking potential time variations of specification in the stochastic discount factor (SDF) model. This paper presents a new method for capturing the time-varying sparsity of factor models by identifying heterogeneous structural breaks instrumented by macroeconomic regimes. Our empirical findings highlight that factor model specification changes over time. We identify time-invariant factors, such as `REG` and `STR`, as well as time-varying factors, such as `IMD`, `BAB`, and `IVOL`, selected in different periods in response to macroeconomic-targeted regime switching. The collective explanatory power of these 20 risk factors is high during periods of high interest rates or low market valuation, but their effectiveness declines when market liquidity is high. Finally, we evaluate factors by modeling unsynchronized factor discovery using unbalanced panel data to account for heterogeneous academic publication timings.

Key Words: conditional asset pricing, heterogeneous structural breaks, macroeconomic regimes, sparsity, time-varying model specifications.

JEL Classification: C14, G11, G12.

*We thank Doron Avramov, Li Chen (discussant), Gustavo Freire, Xiaoyi Han, Caio Ibsen Rodrigues de Almeida, Jingyu He, Dacheng Xiu, Guofu Zhou, and seminar and conference participants at the City University of Hong Kong, Southern University of Science and Technology, Xiamen University, Financial Econometrics Conference at Lancaster University, Econometric Workshop for Time-varying Coefficient Models at Hunan University, Youth Econometrician in Asia-Pacific conference at Fudan University, Exploring the Frontiers of Financial Econometrics in the Big Data Era at KAIST, Sixteenth Annual Risk Management Conference at NUS, 2023 Asian Meeting of the Econometric Society (Singapore), 2023 International Symposium on Advances in Panel Data and Time Series Econometrics, for invaluable comments and discussions.

†Department of Economics and Finance, City University of Hong Kong. Email: liyuan.cui@cityu.edu.hk.

‡Department of Management Sciences, City University of Hong Kong. Email: gavin.feng@cityu.edu.hk.

§Center for Forecasting Science, Chinese Academy of Sciences; School of Economics and Management, University of Chinese Academy of Sciences (UCAS); MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation at UCAS. Email: yh20@cornell.edu and ymhong@amss.as.cn.

¶Department of Economics and Finance, City University of Hong Kong. Email: jiangyang4-c@my.cityu.edu.hk.

1 Introduction

Empirical asset pricing employs linear factor models to approximate the Stochastic Discount Factor (SDF) and explain cross-sectional returns. However, the accumulating "factor zoo" challenge further complicates the factor model specification and selection (Cochrane, 2011). Many empirical studies evaluate and select factors solely based on their historical aggregate performance, implicitly assuming a time-invariant model specification or sparse factor structure, and overlooking potential time variations in factor composition in the SDF. This paper aims to investigate the following research question: *Does the factor model specification change over time?*

Though the time-varying modeling allows changes in factor risk exposures or premia (Bollerslev, Engle, and Wooldridge, 1988; Ferson and Harvey, 1999; Nagel and Singleton, 2011; Roussanov, 2014; Gagliardini, Ossola, and Scaillet, 2016; Smith and Timmermann, 2021), the *illusion of sparsity* (Giannone, Lenza, and Primiceri, 2021) still dominates the empirical literature in the search for the small but fixed set of factors for the entire history of observations. However, there is evidence of *time-varying sparsity* in factor performance, indicating that the active set of factors changes over time. For example, Figure 1 presents changes in the five-year significance of risk premia (t-statistics proportional to Sharpe ratios) for several well-known risk factors.

These *heterogeneous structural breaks* in factor performance can be caused by their unsynchronized discoveries. One possible economic explanation is post-publication bias (McLean and Pontiff, 2016), which may cause early-discovered useful factors to disappear as risk premia decrease due to market efficiency. If one wants to assess the marginal contribution of factors to the SDF while controlling for others, a time-varying model specification is more appropriate than a fixed one when considering factor selection in the accumulating zoo of factors. Considering the increasing number of control factors, Feng, Giglio, and Xiu (2020) find that new ones may emerge to complement or replace old ones, leading to factor proliferation.

This paper proposes a new perspective on time variation by highlighting the potential changes in the factor model specification over time, which achieves time-varying sparsity in variable selection. The time-varying sparsity can also be motivated as a special case of a model with heterogeneous structural breaks. For example, Smith and Timmermann (2022) differentiate between market-wide and style-specific breaks in the panel of stock returns because certain industry or characteristic-sorted portfolios are affected heterogeneously. We present a sparse fused GMM approach (SFGMM) for estimating a time-varying coefficient model, enabling the time-varying model specifications. Our method provides an alternative estimation for the dynamic SDF model, instrumented by a high-dimensional set of aggregate predictors and test assets. This allows for selecting different factors and their varying marginal contributions to the SDF in various macroeconomic regimes, because investors respond to business cycles and macroeconomic conditions (Nagel and Singleton, 2011).

The goal of SFGMM is to consistently identify time-varying sets of useful factors whose risk price parameter values may experience structural breaks at different time periods. Rather than risking model misspecification errors in modeling the function of time variation, SFGMM estimates the entire $T \times p$ parameter values for the time-varying coefficients of p factors in a total of T periods. To incorporate the time-varying patterns in Figure 1, SFGMM promotes similarities of parameter values within each stable regime while enhancing the goodness of fit to the cross-sectional returns by allowing for the emergence of different regimes. Therefore, our SFGMM method can detect heterogeneous structural breaks for risk factors and estimate their SDF loadings between the breaks simultaneously.

Our SFGMM utilizes the generalized method of moments (GMM), a versatile framework encompassing various estimation and inference methods (Hansen, 1982), widely applied in estimating the SDF. The objective function includes a quadratic form of sample moment vectors, a smoothness condition, and a penalty term for model complexity. We introduce a double regularization scheme to identify heterogeneous structural breaks in the roles of risk factors within the SDF, utilizing the high dimension of instrumental variables (IV) and a large cross section of test assets. In a nutshell, our proposed method has two appealing properties that extend beyond applications to asset pricing: (i) The time-varying sparsity implies time-varying model specifications, unlike the global sparsity that restricts a fixed model selection over time. (ii) The heterogeneous structural breaks produce time-varying coefficient values, different from the constant-coefficient model. Therefore, the proposed method can capture the time variation for both presence and magnitude of a factor's marginal explanatory power, while controlling the accumulation of other factors.

Empirical Highlights. To investigate the empirical facts of time-varying sparsity, we study 20 widely cited U.S. factors to estimate the dynamic composition of the SDF from 1972 to 2021. We employ the SFGMM to detect heterogeneous structural breaks in multiple risk factors. For example, previously strong factors, such as short-term reversal (`STR`) and idiosyncratic volatility (`IVOL`), have been deselected due to weak performance in the 2010s. In contrast, newly published factors like betting-against-beta (`BAB`) and expected growth (`REG`) consistently exhibit positive risk prices. Across macroeconomic-targeted regimes, our results identify time-invariant factors, such as `REG` and `STR`, as well as time-varying factors, such as `IMD`, `BAB`, and `IVOL`, that have different selected models in response to regime switching. These findings confirm that there have been changes in the factor model specification over time.

Second, we find that the collective explanatory power of these 20 risk factors is high during periods of high interest rates or low market valuation, but their effectiveness declines when market liquidity is high. Timing on changing macroeconomic regimes, the better-fitted time-varying specified factor model provides superior investment performance. The restricted SFGMM approach achieves an annualized Sharpe ratio of 2.59 and a monthly alpha of 1% for the in-sample analysis, both of which are the highest among long-only investments. Moreover, the restricted SFGMM strategy exhibits an annualized maximum drawdown of only 6.15%. These performance results underscore the effectiveness of our approach in fitting the return dynamics and generating superior investment performance as the SDF.

Finally, many empirical studies assume the simultaneous discovery of all factors for balanced panel data modeling. However, SFGMM's ability to model time variation allows it to handle unbalanced panel data and unsynchronized factor discovery. For example, profitability and investment factors demonstrate positive average returns but are not selected in our post-publication analysis. This doesn't necessarily mean these factors vanish after publication but suggests they may not make an incremental contribution. These empirical findings highlight that factor model specification changes over time, and our SFGMM provides a useful estimation procedure to capture the time-varying sparsity of asset pricing factor models.

Literature Position in Asset Pricing. Numerous empirical evidence suggests structural changes in asset pricing models (Pástor and Stambaugh, 2001; Bekaert et al., 2002; Ang and Bekaert, 2002; Smith and Timmermann, 2021, 2022). First, the factor risk price or SDF weights may change due to macroeconomic conditions (e.g., Nagel and Singleton, 2011). Second, the time variations and dynamics in factor risk premia are well-studied (e.g., Zhou, 1994; Avramov and Chordia, 2006; Gagliardini et al., 2016). Third, linear factor models commonly use time-varying factor loadings or betas (e.g., Bollerslev et al., 1988; Ferson and Harvey, 1999; Lewellen and Nagel, 2006). Whereas conditional asset pricing has utilized various time-varying coefficient models (e.g., Li and Yang, 2011; Ang and Kristensen, 2012; Adrian et al., 2015; Connor et al., 2012), the time-varying specified model remains underdeveloped. Our paper presents the first approach considering unknown heterogeneous structural breaks within a time-varying coefficient model. Our SFGMM approach is similar to Roussanov (2014) and Cui, Feng, and Hong (2022) in evaluating conditional implications of asset pricing models without specific parametric structures.

In the empirical finance literature, sparse modeling (e.g., Chinco et al., 2019; Freyberger et al., 2020) has become a popular solution to the high-dimension of factors or characteristics. Recent factor selection studies aim to maximize either asset pricing model fitness (e.g., Bryzgalova et al., 2023), factor investment performance (e.g., Barillas and Shanken, 2017), or both (e.g., Feng et al., 2023). Whereas previous studies have mainly examined factors' historical aggregated performance for modeling time-invariant sparsity, our approach allows for time-varying sparsity, which is examined with evidence of unsynchronized factor discovery in academic publications. Bayesian studies usually select factors by averaging their explanatory power (Barillas and Shanken, 2018; Avramov et al., 2023; Bryzgalova et al., 2023), but do not explicitly account for the time variation of factor models.

Contrary to conventional wisdom, Kelly, Malamud, and Zhou (2023) argue against the illusion of simple sparse models and prove that these models severely understate return predictability compared to "complex" models. Additionally, Didisheim, Ke, Kelly, and Malamud (2023) find that the best factor model combines tens of thousands of factors. In addition, Cong, Feng, He, and Li (2022) examines uncommon factor selections in cross-sectional and time-series data using a tree-based clustering method.

Methodology Innovations. Lasso-type regularizations have been widely used for variable selection and classification in linear regression models (Tibshirani, 1996; Tibshirani et al., 2005). A recent paper by Giannone, Lenza, and Primiceri (2021) documents the "illusion of sparsity" in many panel datasets, including determinants of economic growth, asset pricing factors, and the decline in crime rates. One more realistic but challenging modeling task is to allow for time-varying sparse model specifications. For modeling time-varying coefficients, our paper presents a solution for detecting heterogeneous structural breaks by considering the potential "time-varying sparsity", which is unknown and requires consistent estimation.

SFGMM achieves consistent time-varying model selection by leveraging sparsity structures in two dimensions. By applying a fused Lasso penalty, we estimate unknown structural changes and encourage similarity in coefficient values between consecutive periods, resulting in a sparse feature of heterogeneous structural breaks. Additionally, we utilize heterogeneous Lasso penalties to account for the individual duration of structural changes, leading to time-varying sets of effective covariates with nonzero risk prices. Figure 2 displays the resulting time-varying risk price estimates for 20 risk factors, which exhibit heterogeneous structural changes in different periods. By relaxing the concept of sparsity, the interpretation of time-varying variable selection is enhanced, which enhances the potential applications of the time-varying coefficient model. This represents one of the primary contributions of our paper to the literature on time-varying coefficient models, asset pricing, and beyond.

Our SFGMM approach, similar to Cui, Feng, and Hong (2022), treats time-varying coefficient values as a high-dimensional parameter vector without assuming their relationship with covariates or making distributional assumptions. However, our approach differs in several key aspects: (i) We incorporate time-varying variable selection, ensuring model selection consistency even with a high dimension of covariates. (ii) We propose a novel double-penalization scheme to address parameter-value oscillations, enable time-varying sparsity, and achieve consistent model selection. It combines the fused Lasso and weighted Lasso penalties (Tibshirani et al., 2005) on time variation. (iii) We establish large sample properties for high-dimensional dependent data under milder conditions. (iv) Our approach is particularly valuable for empirical practitioners as it allows for time-varying variable selection and enhances the understanding of underlying economic mechanisms.

Literature Position in Structural Breaks. Statistical tests for structural breaks are developed in various modeling procedures, such as linear regressions with endogeneity, high-dimensional modeling, factor analysis, and GMM (Andrews, 1993; Stock and Watson, 1996; Bai and Perron, 1998; Corradi and Swanson, 2014; Baltagi et al., 2021). There is a vast literature on estimating structural breaks. The traditional approach has assumed prior knowledge of breakpoints, including their positions and numbers, which apply to all covariates, known as common structural breaks (Bai, 2010; Baltagi et al., 2016). Lasso regularization is used to estimate unknown breaks through a multi-stage process in recent literature. Chan et al. (2014) propose a two-step Lasso procedure, Cheng et al. (2016) propose an adaptive group-Lasso estimator for determining the number of factors for a single change point, and Safikhani and Shojaie

(2022) proposes a three-stage procedure for simultaneous estimating change points and parameters of high-dimensional piecewise vector autoregressive models.

Our SFGMM method differs from existing methods by allowing for heterogeneity in the timing of structural breaks for different covariates and letting the sets of effective covariates with nonzero coefficients vary over time through a global one-step procedure. The innovation is to treat the time-varying parameter values for all potential covariates as a high-dimensional parameter vector. The large dimensionality of test assets and IVs becomes a "blessing" for our SFGMM method because it enables the construction of sufficient moment conditions to consistently estimate the high-dimensional parameter vector and achieve model selection consistency. Beyond asset pricing, SGM provides a general framework for estimating heterogeneous structural breaks in panel data modeling.

This paper is structured as follows: In Section 2, we introduce the challenge of estimating the dynamic SDF model, our SFGMM estimator, and its statistical properties. Section 3 analyzes the time-varying factor model specification of the SDF, and Section 4 compares model performances under different macroeconomic regimes. We illustrate the factor proliferation study in Section 5, and conclude the paper in Section 6. The appendix presents the simulation study and SFGMM's assumptions and global properties. We provide the remaining mathematical details in the online appendix.

2 Time-Varying Specified Factor Model

2.1 Dynamic Stochastic Discount Factor Models

In the absence of arbitrage, a time-varying stochastic discount factor model, m_{t+1}, exists such that for any traded asset i with an excess return at time t of r_t, we have the conditional moment equation:

$$E[m_{t+1}r_{t+1}|I_t] = \mathbf{0}_K, \tag{1}$$

where r_{t+1} is an $K \times 1$ vector of excess returns on K assets, and I_t is the economist's information set whose complexity grows with time. A beta pricing model can be cast in the SDF framework by specifying the SDF as a linear function of f_{t+1}, where f_{t+1} is a $p \times 1$ vector of risk factors:

$$m_{t+1} = 1 - \gamma'_{o,t+1} f_{t+1}, \tag{2}$$

where $\gamma_{o,t+1}$ is a $p \times 1$ vector of time-varying SDF weights or risk price parameters, the number of factors p can be large.

The parameter $\gamma_{o,t+1}$ (or γ_o) is critical as a risk price coefficient in the SDF model's conditional asset pricing literature. According to Chapter 13 of Cochrane (2009), risk premium compensates for holding risk and is typically positive, but risk price differs. To price the cross section, it is more useful to examine a factor's SDF loading than its risk premium. Risk premia should be positive, but risk prices can be negative when combined with other factor exposures in the SDF model. For distinguishing useful factors from redundant or useless ones, risk price assesses a factor's marginal explanatory power. Therefore, estimating the time-varying risk price coefficient values using SFGMM can provide valuable insights into conditional asset pricing models.

A closed-form solution exists for the unconditional model when $\gamma_{o,t+1} = \gamma_o$ is constant. In this case, GMM estimation is equivalent to estimating a cross-sectional regression with expected returns on factor-return covariance (Cochrane, 2009). Prior studies have focused on estimating a sparse and constant γ_o for the cross-sectional regression (Feng et al., 2020). Variable selection becomes simpler when focusing on risk prices rather than risk premia, as the regressors are univariate factor-return covariances rather than multivariate betas. However, if the structure of $\gamma_{o,t+1}$ is potentially time-varying and sparse due to heterogeneous structural breaks, the cross-sectional regression may not be feasible. One needs a new methodology to understand better the time-varying presence and magnitude of a factor's marginal explanatory power.

We propose a novel time-varying factor selection method that estimates a factor's risk price $\gamma_{o,t+1}$ without assuming a prior knowledge of structural changes. Our method estimates all realization values of time-varying parameters over the sample period. As discussed in Chapter 10 of Cochrane (2009), conditional modeling uses IVs for GMM estimation. We transform (1) from a conditional SDF representation to an unconditional one by scaling IVs. Our method does not assume the positions of structural changes as a priori, allowing for elements of $\gamma_{o,t+1}$ to be zero in some periods and nonzero in others. Then we have

$$\mathbf{0}_{\tilde{r}} = E(u_{t+1} \otimes \tilde{z}_t), \tag{3}$$

where $u_{t+1} = (1 - \gamma'_{o,t+1} f_{t+1}) r_{t+1}$ is the $K \times 1$ vector of pricing errors, which is also the ex-post discounted return in asset pricing. $\tilde{z}_t$ is the $\tilde{l}$ vector of IVs, which can be measurable functions of conditioning variables $z_t \in R^l$ from the economist's information set I_t.[1] Economists choose conditioning variables uncorrelated with future pricing errors but weakly correlated with future returns, like lagged returns, predictable firm characteristics, and aggregate predictors.

2.2 Challenges in Empirical Methods

Existing studies often assume a fixed set of ad hoc risk factors and then model their time-varying risk premia or prices (e.g., Connor et al., 2012; Ang and Kristensen, 2012). Choosing a common set of useful factors from an increasingly available set of candidate factors is challenging and can easily suffer from various inference issues, such as weak factors. However, the factor selection literature mainly focuses on using a fixed set of factors to maximize asset pricing or investment performance(e.g., Feng et al., 2020; Barillas and Shanken, 2017), ignoring potential structural changes in the presence and magnitude of factors' time-varying roles within the SDF model

Existing studies also assume that structural breaks on covariate parameter values occur synchronously (e.g., Bekaert et al., 2002; Smith and Timmermann, 2021), which

[1] The number of moment conditions $\tilde{r} = K\tilde{l}$. Given that we use the k-fold repeated cross-validation (CV) method to select tuning parameters, we denote the effective number of moment conditions used in the estimation procedure as $r = \tilde{r}(k-1)/k$, and the notation becomes consistent with our general model setup in (4).

rules out a more realistic situation where different factors may experience structural breaks at different time points. Though Smith and Timmermann (2022) allows for heterogeneous structural breaks, their method does not cover the GMM framework. Unfortunately, such difficulties are inevitably escalated in high-dimensional settings. Our SFGMM aims to consistently estimate heterogeneous structural breaks while simultaneously achieving time-varying factor selection.

Modeling time-varying coefficients as functions of conditioning variables is a popular choice (e.g., Avramov and Chordia, 2006; Nagel and Singleton, 2011; Adrian et al., 2015; Kelly et al., 2019), but there might be a model specification issue when selecting or shrinking the potentially long list of conditioning variables. Unlike traditional variable selection methods, our SFGMM estimates all time-varying parameters using a comprehensive set of moment restrictions without involving selecting or shrinking conditioning variables. Our method can effectively handle large datasets and mitigate potential issues caused by model misspecification. In essence, the high-dimensional nature of the asset pricing problem proves beneficial and assists in estimating the time-varying SDF weights and detecting heterogeneous structural breaks.

2.3 Sparse Fused GMM Estimator

We generalize the problem in (3) and consider the following general econometric model, which has a unique true $T \times p$ parameter matrix $\Gamma_o = (\gamma_{o,1}, \cdots, \gamma_{o,T})'$ satisfying the moment condition

$$E\big[e(U_t, \gamma_{o,t})\big] = \mathbf{0}_r, \tag{4}$$

where E denotes the mathematical expectation, $e(U_t, \gamma_t)$ is a $r \times 1$ residual vector, U_t contains observable variables, and $\gamma_t = (\gamma_{t,1}, \cdots, \gamma_{t,p})'$ is a $p \times 1$ vector of parameters that may vary over time t. Let $\Gamma = (\gamma_1, \cdots, \gamma_T)'$ be the $T \times p$ parameter matrix in a compact subset $S \in R^{T \times p}$. We allow for many candidate covariates whose dimension p can be large, and coefficient values may be zero for certain periods but nonzero for others. Thus, for identification purposes, we require $r \geq pT$, that the number of moments is at least as many as the total number of parameters.

We summarize the time-varying parameter structure with some notations as follows. For the jth candidate covariate with $j \in \{1, \cdots, p\}$, the coefficient vector $\gamma_{\cdot,j} = (\gamma_{1,j}, \cdots, \gamma_{T,j})'$ can be classified into N_j sets $G_1^j, \cdots, G_{N_j}^j$, where $G_{n_1}^j \cup G_{n_2}^j = \emptyset$ for any $n_1 \neq n_2$, and $\cup_{n=1}^{N_j} G_n^j = \{1, \cdots, T\}$. We let $N = \sum_{j=1}^{p} N_j$ be the total number of common values in all regimes and $N_0 = 0$. We further denote $\theta = (\theta_1, \cdots, \theta_N) \in R^N$ and $|\mathbb{G}| = \left\{|G_n^j|, j \in \{1, \cdots, p\}, n \in \{1, \cdots, N_j\}\right\} = \{|G_1|, \cdots, |G_N|\}$, where $|G_{n'}| = |G_n^j|$ and $\theta_{n'} = \theta_n^j$ denote the duration and common parameter value for the jth covariate in its nth regime G_n^j for $n' = \sum_{m=0}^{j-1} N_m + n$. There exists a selection matrix M, which reflects the information on structural breaks and bridges the parameter matrix and common value vector that $vec(\Gamma) = M\theta$.[2]

We consider a sparsity feature in θ to allow for time-varying sets of active covariates. We can partition the common value parameter vector into two subvectors $\theta = (\theta_{\mathbb{A}}', \theta_{\mathbb{Z}}')'$, where $\theta_{\mathbb{A}} \in R^q$ and $\theta_{\mathbb{Z}} \in R^{N-q}$ contain the nonzero and zero elements of θ, respectively. To save notations, we use $e(U_t, \gamma_t) = e(U_t, \theta)$ exchangeably in moment conditions in the rest of the paper.

We allow the following general time-varying structures for a large pool of candidate covariates. First, when $N_j = 1$, the jth covariate exhibits a homogeneous impact in the econometric model (4), whose coefficient values share a constant common value θ_1^j over time $\gamma_{t,j} = \gamma_{\tau,j} = \theta_1^j$ for all $1 \leq t, \tau \leq T$ and $G_1^j = \{1, \cdots, T\}$. We allow zero or nonzero θ_1^j, whereas a nonzero value makes the jth covariate useful and should thus be selected in our model; otherwise, such a covariate is treated as redundant.

Second, for the jth candidate covariate, when $N_j > 1$, $N_j - 1$ structural breaks exist in the jth covariate's parameter values with $\gamma_{t,j} = \theta_n^j = \theta_{n'}$, which is the common coefficient value that the jth covariate shall take when the time period t belongs to the nth time regime, $t \in G_n^j$, for $n \in \{1, \cdots, N_j\}$ and $n' = \sum_{m=0}^{j-1} N_m + n \in \{1, \cdots, N\}$. Importantly, we allow for time-varying model selection that the jth covariate may

[2] $vec(\Gamma) = (\gamma_1', \cdots, \gamma_T')'$ and $M = (M_1', \cdots, M_T')'$ is the $pT \times N$ selection matrix with M_t being a $p \times N$ matrix with the (j, k')th element $M_{t,j,k'} = 1$ for $t \in G_k^j$, $k' = \sum_{m=0}^{j-1} N_m + k$, and $M_{t,j,k'} = 0$ otherwise for all $t \in \{1, \cdots, T\}$.

have zero parameter values in certain periods but exhibit nonzero impacts in others.

Our SFGMM estimator proceeds in situations where the economist can or can not acquire complete information on structural breaks, respectively, as follows:

1. Prior Information on the Timing of Structural Breaks

   Given the information on structural breaks, economists can construct the selection matrix M that $vec(\Gamma_o) = M\theta_o$; thus we only need to estimate θ_o through the following penalized GMM procedure:

$$\tilde{\theta} = \arg\min_{\theta \in R^N} \frac{1}{r} \|g_T(\theta)\|^2 + \sum_{n=1}^{N} |G_n| * P_\eta(|\theta_n|), \tag{5}$$

   where $\|\cdot\|$ is the l_2 norm of a vector, $g_T(\theta) = T^{-1}\sum_{t=1}^{T} e(U_t, \gamma_t)$ is the sample analog of the vector residual, and the quadratic term is normalized by r to offset the impact of the divergent dimension of moment restrictions.

   We employ the duration of each time region $|G_n|$ as the heterogeneous penalty weights on different parameter values' importance in the model. To encourage time-varying model recovery, we further shrink individual parameter values via the l_1-type penalty function $P_\eta(|\cdot|)$.[3]

2. Time-Varying Sparsity and Unknown Structural Breaks

   When economists are blind to structural changes, we propose obtaining $\hat{\Gamma}$ directly from the following general double-penalized SFGMM procedure:

$$\hat{\Gamma} = \arg\min_{\Gamma \in S} \frac{1}{r} \|g_T(\Gamma)\|^2 + \sum_{j=1}^{p}\sum_{t=2}^{T} P_\lambda(|\gamma_{t,j} - \gamma_{t-1,j}|) + \sum_{j=1}^{p}\sum_{t=1}^{T} P_\eta(|\gamma_{t,j}|), \tag{6}$$

   where the fused Lasso regularization $\sum_{j=1}^{p}\sum_{t=2}^{T} P_\lambda(|\gamma_{t,j} - \gamma_{t-1,j}|)$ controls for the total amount of structural breaks in the sampling period by keeping a few

[3] Examples of penalty functions, $P_\eta(\cdot)$ and $P_\lambda(\cdot)$, include the smoothly clipped absolute deviation (SCAD) penalty as in Fan and Li (2001) that $P'_\lambda(x) = \lambda\{\mathbb{1}(x \le \lambda) + (a\lambda - x)_+/((a-1)\lambda)\mathbb{1}(x > \lambda)\}$ and the minimax concave penalty (MCP) in Zhang (2010) with $P'_\lambda(x) = sign(x)[\lambda - |x|/C]$ if $|x| \le C\lambda$ and 0 otherwise. The Lasso penalty in Tibshirani (1996) applies under the irrepresentable requirement.

nonzero piece-wise constant parameter values over time while forcing the differences between coefficients in most periods to be precisely zero for each covariate. The second Lasso penalty, $\sum_{j=1}^{p}\sum_{t=1}^{T} P_\eta(|\gamma_{t,j}|)$, aims at time-varying sparsity recovery by shrinking individual parameter values.[4]

We recover the positions and magnitudes of structural breaks in parameter values by balancing the goodness of fit and encouraging sparsity in parameter differences. We elaborate on the selection of tuning parameters η and λ and the implementation algorithm for (5) and (6) in the appendix.

2.4 Statistical Inference

We provide large sample properties of our SFGMM estimator with and without prior knowledge of structural changes, respectively. Moreover, we also derive the limiting distribution of the estimators under a jointly large r, p, and T design. In Appendix B, we provide primitive conditions for the wide applicability of our method and prove the following theorems.

Theorem 1 (Consistency)**.** *Under Assumptions A.1-A.5 in Appendix B. A strict local minimizer $\tilde{\theta} = (\tilde{\theta}'_{\mathbb{A}}, \tilde{\theta}'_{\mathbb{Z}})'$ of the penalized GMM criterion (5) exists such that*

$$\|\tilde{\theta}_{\mathbb{A}} - \theta_{o\mathbb{A}}\| = O_p\big(\max\{\sqrt{q\log r/T}, \sqrt{q}\{\log[r(T+1)]\}^{1/\bar{\nu}}/T\}\big).$$

Moreover, with probability tending toward one as $T \to \infty$, we have $\tilde{\theta}_{\mathbb{Z}} = \mathbf{0}$ as $T \to \infty$.

Theorem 2 (Selection Consistency)**.** *Under Assumptions A.1-A.5 in Appendix B. Then,*

$$P(\tilde{\mathbb{I}} = \mathbb{I}) = 1,$$

where $\mathbb{I}$ is the support of the indexes of the nonzero components $\mathbb{I} = \{n : 1 \le n \le N, \theta_{o\mathbb{A},n} \neq 0\}$ and $\tilde{\mathbb{I}} = \{n : 1 \le n \le N, \tilde{\theta}_n \neq 0\}$.

[4]In the empirical implementation, one can insert a diagonal matrix to allow different penalty levels of different factors with prior domain knowledge. For example, one can assign time-varying penalty weights for the performance of risk factors to utilize information.

Theorem 3 (Asymptotic Normality). *Under Assumptions A.1-A.5 in Appendix B and assume* $P'_\eta(d_T) = o\big(1/(\sqrt{qT}|G_{\max}|)\big)$ *with* $d_T = \min_{1\le n\le q}|\theta_{o\mathbb{A},n}|/2$ *as half of the minimum signal. We define* $\Sigma_a = a'[\psi_{\mathbb{A}T}\psi'_{\mathbb{A}T}]^{-1}\psi_{\mathbb{A}T}V\psi'_{\mathbb{A}T}[\psi_{\mathbb{A}T}\psi'_{\mathbb{A}T}]^{-1}a$ *for some unit vector* $a \in R^q$ *with* $\|a\|^2 = 1$, $V = var\big(\sqrt{1/T}\sum_{t=1}^T e(U_t, \gamma_{o,t})\big)$, *and* $\psi_{\mathbb{A}T} = T^{-1}\sum_{t=1}^T E\partial e(U_t, \gamma_{o,t})/\partial\theta_{\mathbb{A}}$. *Then, as* $T \to \infty$, *we have*

$$\Sigma_a^{-1/2}\sqrt{T}a'(\tilde{\theta}_{\mathbb{A}} - \theta_{o\mathbb{A}}) \xrightarrow{d} N(0,1).$$

For the estimation consistency result, Theorem 1 implies that when the time-varying structure is known as a prior, $\tilde{\theta}$ that solves the penalized GMM criterion (5) converges to the true common values θ_o with probability tending toward one. In general, the convergence rate of the nonzero elements contains two parts. The first term demonstrates the trade-off when we employ many moment conditions to avoid specifying the data-generating processes for time-varying parameters. The second term reflects the influence of data's serial dependency and thinness of the moment condition tails, measured by $\bar{\nu}$, because they affect how fast sample analogs could converge in probability to their population moments.

For the selection consistency result, Theorem 2 confirms that even though the economist may face a large pool of potential covariates, he/she can still detect useful factors from redundant ones and estimate their time-varying influences simultaneously. For the asymptotic normality result, Theorem 3 establishes the limiting distribution using a self-normalized representation for the non-zero values of our SFGMM estimator when the time-varying structure is known.

Theorem 4. *Under Assumptions A.1-A.5 in Appendix B. A positive constant* c *exists such that* $\min_{1\le n\le q}|\theta_{o\mathbb{A},n}| > c\eta$. *Let* $b_T = \min_{1\le j\le p, N_j>1}\min_{n_1\ne n_2}|\theta^j_{on_1} - \theta^j_{on_2}|$ *and* $b_T > c\lambda$, *with* $\lambda \gg 1/\sqrt{T}$. *A local minimizer* $\hat{\Gamma}$ *of the penalized SFGMM criterion (6) exists such that*

$$P(\hat{\Gamma} = \tilde{\Gamma}^*) = 1,$$

where $\tilde{\Gamma}^*$ *is the oracle estimator corresponding to* $\tilde{\theta}^* = (\tilde{\theta}^{*\prime}_A, \mathbf{0}')'$ *that* $vec(\tilde{\Gamma}^*) = M\tilde{\theta}^*$, *which* *minimizes the following moment conditions:* $\tilde{\theta}^*_{\mathbb{A}} = \arg\min_{\theta_{\mathbb{A}} \in R^q} \frac{1}{r} \|g_T((\theta'_{\mathbb{A}}, \mathbf{0}')')\|^2$.

Theorem 5 (Asymptotic Variance Estimation)**.** *Under Assumptions A.1-A.5 in Appendix B. $\hat{\theta}$ corresponds to $\hat{\Gamma}$, a local minimizer of the penalized SFGMM criterion (6). A positive constant c_8 exists such that $E|e_j(U_t, \gamma_{o,t})|^{4+2\delta} < c_8 < \infty$ for all $1 \le j \le r$ and $1 \le t \le T$. Suppose $m_T \max_{1 \le t \le T} \|\hat{\gamma}_t - \gamma_{o,t}\| = o_p(1)$ and $m_T/\sqrt{T} = o(1)$. For some unit vector $a \in R^q$ with $\|a\| = 1$, define $\hat{\Sigma}_a = a'[\hat{\psi}_T \hat{\psi}'_T]^{-1} \hat{\psi}_T \hat{V} \hat{\psi}'_T [\hat{\psi}_T \hat{\psi}'_T]^{-1} a$, where $\hat{\psi}_T \equiv T^{-1} \sum_{t=1}^T \partial e(U_t, \hat{\Gamma})/\partial \theta_{\mathbb{A}}$, $\hat{V} = \sum_{j=-(T-1)}^{T-1} k(j/m_T) \hat{\Omega}(j)$, and*[5]

$$\hat{\Omega}(j) = \begin{cases} \dfrac{1}{T} \displaystyle\sum_{t=j+1}^{T} e(U_t, \hat{\gamma}_t) e(U_{t-j}, \hat{\gamma}_{t-j})' & \text{for } j \ge 0, \\ \dfrac{1}{T} \displaystyle\sum_{t=-j+1}^{T} e(U_{t+j}, \hat{\gamma}_{t+j}) e(U_t, \hat{\gamma}_t)' & \text{for } j < 0. \end{cases}$$

Then, as $T \to \infty$, we have $\hat{\Sigma}_a - \Sigma_a \xrightarrow{p} 0$, and $\hat{\Sigma}_a^{-1/2} \sqrt{T} a'(\hat{\theta}_{\mathbb{A}} - \theta_{o\mathbb{A}}) \xrightarrow{d} N(0,1)$.

For the selection consistency result, Theorem 4 implies that even though economists may not acquire the structural break information, they can also establish the estimation and variable selection consistency through our SFGMM procedure.

Theorem 5 provides a consistent variance-covariance estimator for our proposed estimators. It employs the well-known class of heteroskedasticity and autocorrelation (HAC) estimators as in Newey and West (1987) and Andrews (1991) without having to estimate observations' serial dependency levels. When a is a unit vector whose nth entry equals one and zero otherwise, Theorem 5 provides the limiting distribution for the nth regime. While a can be flexible unit vectors, it allows us to perform a joint test on the significance of a particular set of parameter values' importance in the model. Hence, empirical practitioners can perform inference and construct confidence intervals in practice. In finite samples, researchers have been concerned about the undesirable coverage probabilities for HAC and the resultant over-rejections in practice.

[5] m_T is a smoothing parameter that grows with the sample size T and $k(\cdot)$ is a real-valued kernel function, which generally declines as j increases, with $k(0) = 1$. The weights $k(j/m_T)$ satisfy $|k(j/m_T)| \le C$ for finite constant C and $\lim_{m_T \to \infty} k(j/m_T) = 1$ for each j. Kernels include the truncated, Bartlett, Parzen, Tukey-Hanning, and Quadratic spectral functions. A suitable choice of $k(\cdot)$, such as the Bartlett kernel $k(u) = 1 - |u|$ for $|u| \le 1$ and 0 otherwise, ensures the semi-positive definiteness property of $\hat{V}$ in finite samples.

In the appendix, we also describe an alternative bootstrap procedure for conducting inference and constructing confidence intervals.

We note that $\tilde{\theta}$ and $\hat{\Gamma}$ that solve SFGMM with and without prior information on structural breaks in (5) and (6), respectively, are all local. We establish the appealing global properties for $\tilde{\theta}$ and $\hat{\Gamma}$ in Theorems A.1 and A.2 of Appendix B. Therefore, in the presence of a large pool of candidate factors whose appearance can be time-varying, SFGMM provides a novel estimation procedure to estimate their time-varying values while capturing time-varying model specifications.

3 Time-Varying Factor Selection

3.1 Data

We examine a monthly sample period from 1972 to 2021. For model estimation, we focus on a subsample spanning the first 35 years, specifically from 1972 to 2006. The subsequent 15 years, from 2007 to 2021, are reserved for out-of-sample evaluation.

Factors. We consider a collection of 20 traded factors. First, we include factors from two popular models, namely, the Fama-French 5 factors (Fama and French, 2015), which consist of excess market factor (`MKTRF`), small-minus-big (`SMB`), high-minus-low (`HML`), robust-minus-weak (`RMW`), and conservative-minus-aggressive (`CMA`), and additional q-factors (Hou et al., 2021), which consist of investment (`IA`), profitability (`ROE`), and expected growth (`REG`). Second, we include a few factors commonly used in the investment industry, such as betting-against-beta (`BAB`, Frazzini and Pedersen, 2014), quality-minus-junk (`QMJ`, Asness et al., 2019), and HML Devil (`HMLM`, Asness and Frazzini, 2013).

We also include other highly cited risk factors: 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 et al., 2003), intermediary capital risk (IMD, He et al., 2017), earning surprise (SUE, Rendleman Jr et al., 1982), and net equity issues (NI, Loughran and Ritter, 1995).

Test Assets. We employ a dataset of 285 bivariate-sorted portfolios of U.S. equities.The portfolios include the following categories: 25 sorted by size and book-to-market ratio, 25 by size and operating profitability, 25 by size and investment, and 25 by size and momentum. Additionally, we include 25 portfolios sorted by size and beta, 25 by size and short-term reversal, 25 by size and long-term reversal, 25 by size and return variance, 25 by size and residual variance, 35 by size and net issuance, and 25 by size and accruals.

Instrumental Variables. We include ten equity predictors as conditioning variables $z_t \in R^{10}$, including the 3-month treasury bill rate, inflation, TERM and default factors yields, and various aggregate equity market characteristics (earnings-to-price ratio, stock variance, net equity issues, dividend yields, leverage, and liquidity). We standardize these aggregate macro predictors into $[0, 1]$ by their empirical quantile values in the previous 120 months. This approach allows us to evaluate each predictor's macro-target model performance on the same variation scale. We use these ten conditioning variables z_t to construct 65×1 vector of IVs $\tilde{z}_t$ from the one-month lagged predictors, quadratic terms, and interactions. Therefore, these IVs correlate to future returns but are uncorrelated to future pricing errors.

3.2 Time-Varying Factor Selection

Figure 2 presents a heatmap showcasing the factor selection outcomes using all 20 factors. In the color bar, white denotes factors not selected during the respective period, while darker colors indicate larger magnitudes for the risk price estimate. The heatmap demonstrates the time-varying sparsity of the factor zoo in explaining asset returns across the cross section. Despite the historical persistence of these valuable factors, their marginal explanatory power can fluctuate over time due to varying structural breaks. Therefore, the heatmap serves as a visual summary of the dynamic factor selection process, emphasizing the ever-changing nature of factor risk prices.

We can examine the performance of well-known factors by controlling for other factors when considering the SDF. First, the market (`MKTRF`) and market beta (`BETA`) factors are unselected throughout the sample, whereas the size (`SMB`) factor shows slightly positive risk price estimates before the 1990s. Second, the momentum (`UMD`) and short-term reversal (`STR`) factors are consistently positive for the entire training period, though we find a decline for `STR`. Third, some recently published factors, such as betting-against-beta (`BAB`) and expected growth (`REG`), have exhibited highly positive risk price estimates. Like the uncommon factor models in the cross section (Cong et al., 2022), Figure 2 illustrates uncommon factor models in the time series.

Table 1 summarizes the heatmap in Figure 2, which includes the identities of the selected factors, the number of single or multiple selection periods, and their risk price estimates. The table reports the piecewise-constant risk price estimate during each factor's selection period.[6] Our time-varying specified factor model uniquely detects heterogeneous structural breaks for multiple factors. For example, the intermediary capital risk (`IMD`) factor experiences a few breaks, including a slightly negative period in the 1980s, but remains positive for most periods. The idiosyncratic volatility factor (`IVOL`) exhibits sign switches twice, from slightly negative to unselected and negative. The factor model identified from 1993 to 2007 primarily comprises `UMD`, `BAB`, `STR`, `REG`, `IVOL`, and `IMD`, excluding Fama-French three factors (Fama and French, 1993).

Out-of-Sample Performance Positive in-sample results are promising but may not guarantee similar out-of-sample performance. To mitigate the risks of overfitting associated with regularization methods and a high-dimensional set of IVs, out-of-sample evaluations are crucial. These evaluations assess the robustness of the SDF model's performance by examining the investment performance of the SDF strategy. Our approach updates time-varying risk price estimates every five years using data from the

[6]We have applied the asymptotic normality test to these estimates and found some to be significant, which rejects the hypothesis of a time-variant specified factor model or a time-varying coefficient model. This provides positive statistical evidence for the presence of time-varying sparsity.

previous 35 years. Additionally, the sparse factor selection of SFGMM provides valuable insights into the performance of factors over time.

To ensure smoothness in time variation, we update γ_t every five years using the observations in the previous 35 years. The latest estimates are fixed as the estimation for the following five years. This infrequent update framework is adopted based on our in-sample results, which indicate that the factor's estimation and selection do not frequently change over time. Figure 2 illustrates the time-varying estimates of γ_t in the recent 15 years, demonstrating that the out-of-sample γ_t forecasts capture information updates and exhibit time variation. It is worth noting that a factor selected in the training sample may disappear if its performance deteriorates during the test period.

Figure 2 illustrates the updated 35-year out-of-sample factor selection. Overall, the results remain stable, aligning with our expectations. However, there are changes in factor selection during the test period. Notably, `UMD` has lost its marginal explanatory power when controlling for all other factors in the past decade. Previously strong factors, like `STR` and `IVOL`, have been deselected in recent years due to their weak performance during the 2010s. On the other hand, newly published factors, like `BAB` and `REG`, consistently exhibit positive risk prices in the test sample. The positive evidence indicates that the factor model specification changes over time and our SFGMM is a useful estimation procedure. The most recent factor model includes `SMB`, `BAB`, liquidity (`LIQ`), `REG`, `IMD`, `ROE`, and net equity issues (`NI`).

4 Factor Model Comparison

4.1 Asset Pricing Performance

Multiple measures exist to evaluate the goodness of fit for an SDF model. The HJ-distance (Hansen and Jagannathan, 1997) is widely used for GMM estimation. We follow the HJ-R^2 proposed in Cui et al. (2022), which assesses the moment condition fitness and reflects an SDF model's relative asset pricing ability over a benchmark. Given a model M_A with $\hat{m}_{t+1} = 1 - \hat{\gamma}'_{t+1} f_{t+1}$, the ex-post normalized cross-sectional pricing error is

$$e(M_A) = \frac{\frac{1}{T-1}\sum_{t=1}^{T-1} \hat{m}_{t+1} r_{t+1}}{\frac{1}{T-1}\sum_{t=1}^{T-1} \hat{m}_{t+1}}. \tag{7}$$

The aggregated cross-sectional pricing errors with equal weights to each asset follow

$$Q(M_A) = e(M_A)' e(M_A).$$

We can adopt this criterion for both constant and time-varying coefficient models. We define the HJ-R^2 as

$$\text{HJ-}R^2 = 1 - \frac{Q(M_A)}{Q(M_B)}. \tag{8}$$

A higher HJ-R^2 value indicates better goodness of fit, and a positive HJ-R^2 value indicates model M_A outperforms the benchmark model M_B. We use the constant-coefficient CAPM as M_B, which assumes the market factor is solely used to approximate the SDF.

The first column of Table 2 shows the unconditional HJ-R^2 results for the monthly training sample. Notice that our time-varying specified factor model is estimated using all IVs. We also include an all-factor model estimated with and without IVs and other time-varying coefficient models. Additionally, we include ad-hoc selected models such as the Fama-French 3- and 5-factor models (FF3 and FF5), and the q-factor model (q5), all estimated using GMM with the same set of all IVs, for comparison. Our findings demonstrate that these models outperform the CAPM model regarding positive HJ-R^2 values. Moreover, time-varying coefficient models consistently outperform constant-coefficient models regarding asset pricing performance. It is worth noting that the all-factor model outperforms the ad-hoc selected factor models, indicating many risk factors have been historically valuable in empirical asset pricing.

Conditional HJ-R^2 over Macroeconomic Regimes. One may prefer evaluating the estimated factor model based on specific macroeconomic regimes. Adjusting the evaluation criteria allows us to assess conditional asset pricing performance by prioritizing certain macroeconomic regimes. This involves using conditional moment equations that aggregate pricing errors, which are then weighted by standardized macro predictors:

$$e_i(M_A) = \frac{\frac{1}{T-1}\sum_{t=1}^{T-1} \hat{m}_{t+1} r_{t+1} z_{t,i}}{\frac{1}{T-1}\sum_{t=1}^{T-1} \hat{m}_{t+1}},$$

where $z_{t,i}$ represents the ith conditioning variable. Hence, we can analogously define the conditional HJ-R^2: for the ith conditioning variable, HJ$_i$-$R^2 = 1 - Q_i(M_A)/Q_i(M_B)$ with $Q_i(M_A) = e_i(M_A)'e_i(M_A)$ and $Q_i(M_B) = e_i(M_B)'e_i(M_B)$.

The conditional HJ-R^2 measures the aggregate pricing errors of estimated factor models, considering different macroeconomic regimes. For example, when analyzing the influence of inflation on risk factors, the conditional HJ-R^2 calculation is assigned higher weights during periods characterized by high inflation. The standardized macro predictors, scaled between 0 and 1 based on empirical quantiles, prioritize high-inflation periods in the inflation-targeted conditional HJ-R^2. This evaluation assesses the sensitivity of factor models to macroeconomic regime changes, while the unconditional HJ-R^2 assigns equal weights to all periods. Our time-varying specified factor model, estimated with IVs, is analyzed using the conditional HJ-R^2 to assess factor model sensitivity in different macroeconomic regimes, offering new economic insights into time-varying factor selection.

Table 2 presents conditional HJ-R^2 results for macro predictor targets, showing that time-varying coefficient models, particularly SFGMM, generally outperform constant-coefficient models in diverse macroeconomic-targeted regimes. Our findings strongly support time-varying asset pricing factors that respond to macroeconomic-targeted regimes. Additionally, the explanatory power of risk factors changes over time. Collecting these 20 factors demonstrates their high explanatory power during periods of high interest rates (high treasury bill rate) or low market valuation (high earnings-to-price ratio or dividend yield). However, their effectiveness declines when market liquidity is high, and those constant-coefficient factor models barely outperform the CAPM under high market liquidity. While the affine GMM also demonstrates positive results, it tends to overfit the training sample. Section 4.3 assesses the out-of-sample investment performance to evaluate the possibility of overfitting.

Our time-varying specified factor model performs well with policy-targeted macro indicators like interest rates (tbl) and inflation. Hence, the factors we have selected (UMD, BAB, STR, REG, IVOL, and IMD) from the training sample prove valuable in explaining volatile returns amidst high interest rates or inflation. Finally, our findings indicate that the all-factor constant-coefficient model without instrumental variables exhibits lower fitness results in certain macroeconomic regimes. Incorporating instrumental variables enhances effective sample size and improves model performance. Next, we will examine factors' heterogeneous structural breaks instrumented by macroeconomic regimes.

4.2 Model Performance under Macroeconomic Regimes

The time-varying sparsity and heterogeneous structural breaks are instrumented by aggregate return predictors and guided by business cycles. Since these macro variables are standardized by their percentiles of the 10-year rolling window, we can define the high and low states of macroeconomic regimes by using the top and bottom 10%. For instance, during high-inflation periods, the annualized changes in inflation have been in the top 10% high over the past decade. The numbers of observations for each state are reported in Table 4. Our next question is understanding the model's performance under different macroeconomic regimes. First, how do the selected factors change in response to regime switching? Second, how do the explanatory performances change for different anomaly-sorted portfolios?

Time-varying factors. Table 3 reports the top five factors selected by our time-varying specified factor model and their average risk price estimates across different high and low macroeconomic states. First, `REG` is the most robust factor in all high or low states for all macro variables with an incredibly high estimate of risk price. Because the risk price estimate is proportional to the tangency factor portfolio weight, `REG` undoubtedly is the most important source of investment performance across all regimes. `STR` is the second most robust common factor in all high states of macroeconomic regimes, but might be replaced in some low states. IMD is important when TERM factor yield is low, and SMB is important when the market aggregate leverage is low. Our time-varying factors are related to the uncommon factors identified in Cong et al. (2022), which finds different selected factor models over cross section and time series.

Our results identify time-invariant factors, such as REG and STR, as well as time-varying factors, such as IMD, BAB, and IVOL, that have different selected models in response to macroeconomic regime switching. For example, IMD is selected in the high-inflation state but not selected in the low-inflation one. Switching from low to high-interest periods, IMD and SMB replace IVOL and LIQ. Therefore, during high inflation and high interest rate periods, these findings are consistent with the strongly procyclical patterns of the intermediary capital risk factor (He et al., 2017). Multiple structural breaks of IMD in Figure 2 are driven by macroeconomic regime switching.

Explanatory power to anomaly-sorted portfolios. The large set of test assets contains many portfolios formed based on size and each anomaly variable from Fama and French (2016). We investigate how effectively the time-varying specified factor model captures their average returns and whether these anomalies can be dissected. Through two-way sorts of size and anomaly variables, we can assess the model performance for different anomaly-sorted portfolios. Table 4 reports the unconditional HJ-R^2 of our time-varying specified factor model on multiple anomaly-sorted portfolios across different macroeconomic regimes.

We find several interesting patterns for various asset pricing performances for these anomaly-sorted portfolios. First, when market liquidity is high and the explanatory power of factors decreases, the B/M-sorted portfolios become difficult to explain. However, two sets of volatility-sorted portfolios are still explained well. These performance results are relative numbers to the constant coefficient CAPM model. We observe similar patterns when inflation is low, and only portfolios sorted by volatility are well-explained. Second, the explanatory power of portfolios sorted by profitability and beta differ significantly depending on the market's net equity issues. Therefore, the anomalies in profitability and beta can be partially explained. Finally, we find no obvious difference for anomaly-sorted portfolio fitness when the market evaluation (earnings-to-price ratio) and volatility differ.

4.3 Investment Performance

Asset pricing theory emphasizes the equivalence between the mean-variance efficient (MVE) portfolio and the SDF, where the minimum variance of the SDF corresponds to the squared maximal Sharpe ratio of the MVE portfolio (Hansen and Jagannathan, 1991). Thus, the SDF aims to explain the cross section of asset returns and maximize the Sharpe ratio. However, it is crucial to assess the robustness of the SDF model through out-of-sample evaluation, which can be done by analyzing the investment performance of the SDF strategy. This evaluation also helps evaluate the investment information provided by the macro-driven factor rotation. While our empirical analysis primarily focuses on traded factors for constructing the time-varying SDF, SFGMM allows for both traded and nontraded factors, such as consumption.

We assess the investment performance of SFGMM by normalizing the time-varying risk price estimates, which serve as portfolio weights (Figure 2). The SDF model is estimated using data from 1972 to 2006, and its out-of-sample performance is evaluated from 2007 to 2021. The results for restricted (long-only) and unrestricted SDF-implied factor investments, including risk-adjusted investment performance, annualized Sharpe ratio, and monthly Jensen's alpha, are presented in Table 5. We also consider downside risk measures associated with trading the SDF or the tangency portfolio to provide a more comprehensive evaluation. We define the maximum drawdown for any overlapping one-year period as

$$MDD = max_{1 \leq t_1 \leq t_2 \leq T}(Y_{t_1} - Y_{t_2}), \text{ s.t. } |t_2 - t_1| \leq 12, \tag{9}$$

where Y_{t_1} and Y_{t_2} refer to the cumulative log return from month 0 to t_1 and t_2, and the duration between t_1 and t_2 is no longer than 12 months.

Similar to Section 4.1, we compare the performance of our time-varying specified factor model with other constant-coefficient and time-varying coefficient models. Like our SFGMM method, these models use their SDF weights as portfolio weights and are rebalanced monthly. Additionally, we assess the performance of a buy-and-hold strategy for excess market return, the equally weighted, and the MVE investment strategies using all factors.

Over the 35-year in-sample period, the restricted SFGMM strategy demonstrates outstanding performance, achieving the highest annualized Sharpe ratio of 2.59 and a monthly alpha of 1% among all long-only investment strategies. It also exhibits one of the lowest annualized maximum drawdowns of only 6.15%. In comparison, the restricted equally-weighted and MVE portfolios, investing in all factors, experience drawdowns of 14.36% and 10.07%, respectively, during the same period. Moreover, the unrestricted SFGMM strategy outperforms its restricted counterpart across all three metrics.

During the 15-year out-of-sample period, the restricted SFGMM strategy outperforms all other time-varying coefficient models. However, the restricted MVE strategy with all factors demonstrates the best performance before considering transaction costs. Furthermore, using instrumental variables (IVs) helps reduce the maximum drawdown for the all-factor models estimated using GMM. These highly positive results highlight the significant value of the macroeconomic conditioning variables utilized in SFGMM, enhancing the optimization of time-varying SDF portfolios.

5 Factor Proliferation

Unsynchronized Factor Discovery. Many empirical studies that assess risk factors or anomalies often assume that all factors are discovered simultaneously, primarily for the convenience of modeling and utilizing balanced panel data. However, this approach can create an illusion of robust asset pricing or investment performance since it employs an ex-post scheme for factor evaluation. To address this issue, researchers are encouraged to incorporate real-time discovery of risk factors, which provides a more accurate representation of their performance.

The proliferation of factors is partly driven by academic publication requirements, which demand newly discovered factors to offer unique and incremental contributions over existing benchmarks. Consequently, these newly identified factors are expected to showcase superior historical performance, leading to the expansion of the factor zoo. This phenomenon is highlighted in Feng et al. (2020), where they demonstrate the emergence of new factors based on the publication year, further emphasizing the importance of considering the dynamics of factor discovery.

We examine the timing of factor discovery in academic publications by evaluating and estimating newly discovered factors beyond existing ones. The SFGMM approach enables us to flexibly model time variation and handle unbalanced panel data, accommodating unsynchronized factor discovery. In the factor selection process for the SDF, a factor can only be considered after the sample end year studied in the original published paper. Until we reach the point of its discovery in the literature, the risk price estimate path is constrained to be zero. While there may be a publication lag, we designate January immediately following the sample end year as the point of consideration in the published version of this paper. Figure 3 presents the factor names alongside the year axis, indicating when they were included in the SDF's factor selection. For instance, only two factors, `MKTRF` and `BETA`, have been included in the selection for the entire sample since 1972, while `REG` is added in 2019.

In Figure 3, the color white in the color bar indicates that the factor was not selected for that specific period. The color remains white before a factor is considered in the factor selection process. This heatmap visualization enables us to assess the real-time explanatory power of each factor since its discovery, which provides a different perspective from the overall factor comparison. This distinction arises because some recently discovered strong factors are not controlled for during earlier periods. For instance, `MKTRF` demonstrates usefulness for a significant portion of the sample period. The increasing positive performance of `MKTRF` contradicts its overall estimates in Figure 2, but it aligns with the monthly average returns depicted in Figure 1.

The explanatory power of some factors, such as `SMB`, `UMD`, `LIQ`, `STR`, and `IVOL`, which are useful in the 1990s and 2000s, have declined or been replaced by newly published factors such as `QMJ`, `BAB`, and `IMD`. The heatmap in Figure 3 illustrates that the estimated risk prices for these factors are consistent with their monthly average returns shown in Figure 1. In contrast, `NI` has consistently maintained its usefulness since its publication, aligning with its overall estimate in Figure 2. These findings strongly reject the fallacious simplification of time-variant sparsity and underscore the significance of methodological assumptions in financial data applications. The demonstrated flexibility of modeling time variation using SFGMM in this post-publication study contributes to the existing empirical literature on factor evaluation.

Conditional Factor Evaluation. Our study provides novel perspectives on the issue of post-publication bias when evaluating risk factors. Previous studies, such as McLean and Pontiff (2016) and Hou et al. (2020), have primarily focused on examining these factors individually. Their conclusion of declining risk premia can be explained by market efficiency, which shows evidence of heterogeneous structural breaks. In contrast, our SFGMM approach allows us to evaluate their asset pricing explanatory power within the SDF while controlling for existing factor discoveries. Our conditional factor evaluation is about the marginal explanatory power of factors.

Some famous factors, such as profitability and investment, exhibit positive performance in Figure 1, but they are not selected in Figures 2 and 3. These results do not necessarily imply that profitability and investment disappear after publication, but rather that they may not provide an incremental contribution over selected factors. The loss of their marginal explanatory power can be due to the proliferation of new risk factors. From this perspective, our study makes a valuable contribution to the literature by emphasizing the importance of considering the time-varying nature of factor risk pricing and highlighting the flexibility of our SFGMM approach in evaluating the post-publication performance of risk factors.

6 Conclusion

Asset pricing factor models are shown to change over time. The presence and magnitude of a factor's marginal explanatory power are time-varying, while controlling the accumulation of other factors. This paper proposes a method for specifying and estimating a time-varying coefficient model for the stochastic discount factor (SDF) to capture the time-varying sparsity of asset pricing factor models. Our method, Sparse Fused GMM (SFGMM), extends the generalized method of moments (GMM) framework by incorporating a comprehensive set of conditioning variables and test assets. By leveraging heterogeneous structural breaks in the composition of the SDF, our approach aims to provide insights into the time-varying nature of risk factors and their associated risk prices by responding to macroeconomic regime switching. Using SFGMM, we have found positive empirical evidence to support the time-varying specified factor model and have revealed its underlying economic interpretation.

First, SFGMM estimates heterogeneous breaks for multiple risk factors. Previously strong factors, such as `STR` and `IVOL`, have been deselected due to weak performance in the 2010s. In contrast, newly published factors like `BAB` and `REG` consistently demonstrate positive risk prices. Second, our results identify time-invariant factors, such as `REG` and `STR`, as well as time-varying factors, such as `IMD`, `BAB`, and `IVOL`, that have different selected models in response to macroeconomic regime switching. Third, the collective explanatory power of these 20 risk factors is high during periods of high interest rates or low market valuation, but their effectiveness declines when market liquidity is high. Finally, using the SFGMM approach, we evaluate the timing of factor discovery in academic publications and model time variation with unbalanced panel data, accommodating unsynchronized factor discovery. Excluding profitability and investment factors from SFGMM does not mean they disappear after publication, as they still show positive average returns. This implies they may not offer an incremental contribution compared to the selected factors. These empirical findings highlight the importance of a flexible modeling approach to capture the time-varying sparsity of asset pricing factor models.

References

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

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.

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariant matrix estimation. *Econometrica 59*(3), 817–858.

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 821–856.

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

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

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.

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). Predicting stock returns. *Journal of Financial Economics 82*(2), 387–415.

Bai, J. (2010). Common breaks in means and variances for panel data. *Journal of Econometrics 157*(1), 78–92.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 47–78.

Baltagi, B. H., Q. Feng, and C. Kao (2016). Estimation of heterogeneous panels with structural breaks. *Journal of Econometrics 191*(1), 176–195.

Baltagi, B. H., C. Kao, and F. Wang (2021). Estimating and testing high dimensional factor models with multiple structural changes. *Journal of Econometrics 220*(2), 349–365.

Barillas, F. and J. Shanken (2017). Which alpha? *Review of Financial Studies 30*(4), 1316–1338.

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

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

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy 96*(1), 116–131.

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.

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

Chan, N. H., C. Y. Yau, and R.-M. Zhang (2014). Group lasso for structural break time series. *Journal of the American Statistical Association 109*(506), 590–599.

Cheng, X., Z. Liao, and F. Schorfheide (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. *Review of Economic Studies 83*(4), 1511–1543.

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.

Cochrane, J. H. (2009). *Asset pricing: Revised edition*. Princeton university press.

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

Cong, L. W., G. Feng, J. He, and J. Li (2022). Uncommon factors for bayesian asset clusters. Technical report, City University of Hong Kong.

Connor, G., M. Hagmann, and O. Linton (2012). Efficient semiparametric estimation of the fama–french model and extensions. *Econometrica 80*(2), 713–754.

Corradi, V. and N. R. Swanson (2014). Testing for structural stability of factor augmented forecasting models. *Journal of Econometrics 182*(1), 100–118.

Cui, L., G. Feng, and Y. Hong (2022). Regularized GMM for time-varying models with application to asset pricing. Technical report, City University of Hong Kong.

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

Didisheim, A., S. Ke, B. T. Kelly, and S. Malamud (2023). Complexity in factor pricing models. Technical report, Yale University.

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

Fama, E. F. and K. R. French (2016). Dissecting anomalies with a five-factor model. *Review of Financial Studies 29*(1), 69–103.

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

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.

Feng, G., W. Lan, H. Wang, and J. Zhang (2023). Anomaly or risk factor? a stepwise evaluation. Technical report, City University of Hong Kong.

Ferson, W. E. and C. R. Harvey (1999). Conditioning variables and the cross section of stock returns. *Journal of Finance 54*(4), 1325–1360.

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.

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

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

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 1029–1054.

Hansen, L. P. and R. Jagannathan (1997). Assessing specification errors in stochastic discount factor models. *Journal of Finance 52*(2), 557–590.

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., 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 (2020). Replicating anomalies. *Review of Financial Studies 33*(5), 2019–2133.

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.

Kelly, B. T., S. Malamud, and K. Zhou (2023). The virtue of complexity in return prediction. *Journal of Finance, Forthcoming*.

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

Lewellen, J. and S. Nagel (2006). The conditional capm does not explain asset-pricing anomalies. *Journal of Financial Economics 82*(2), 289–314.

Li, Y. and L. Yang (2011). Testing conditional factor models: A nonparametric approach. *Journal of Empirical Finance 18*(5), 972–992.

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

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

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

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

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

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

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.

Roussanov, N. (2014). Composition of wealth, conditioning information, and the cross-section of stock returns. *Journal of Financial Economics 111*(2), 352–380.

Safikhani, A. and A. Shojaie (2022). Joint structural break detection and parameter estimation in high-dimensional nonstationary var models. *Journal of the American Statistical Association 117*(537), 251–264.

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 (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics 14*(1), 11–30.

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

Tibshirani, R., M. Saunders, S. Rosset, J. Zhu, and K. Knight (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67*(1), 91–108.

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics 38*(2), 894–942.

Zhou, G. (1994). Analytical GMM tests: Asset pricing with time-varying risk premiums. *Review of Financial Studies 7*(4), 687–709.

Figure 1: Factor Performance for Every Five Years

Notes: This figure displays a heatmap illustrating the time-varying significance of risk premia for 20 risk factors (described in Section 3.1) based on t-statistics of their monthly average returns and annualized Sharpe ratios (in the square brackets) for every five years from 1972 to 2021. We have set the color white to indicate insignificance at the 10% level.

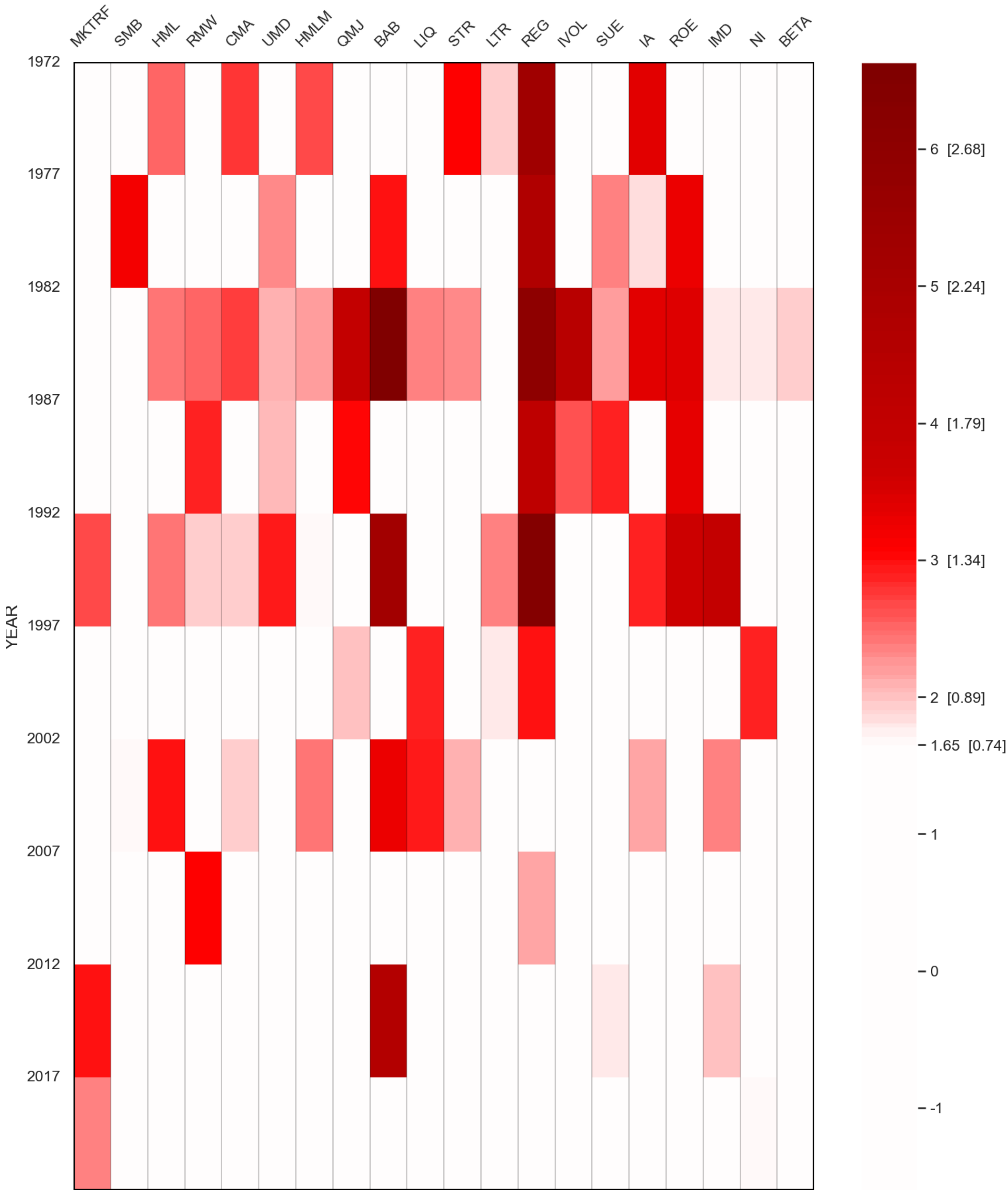


Figure 2: Time-Varying Estimates for Selected Factor Risk Prices

Note: This figure displays a heatmap presenting the factor selection results using all twenty factors (described in Section 3.1). In the color bar, white indicates that the factor is not selected for this period, while a darker color indicates a larger magnitude for the risk price estimate. The training period is from 1972 to 2006, and the test period is from 2007 to 2021.

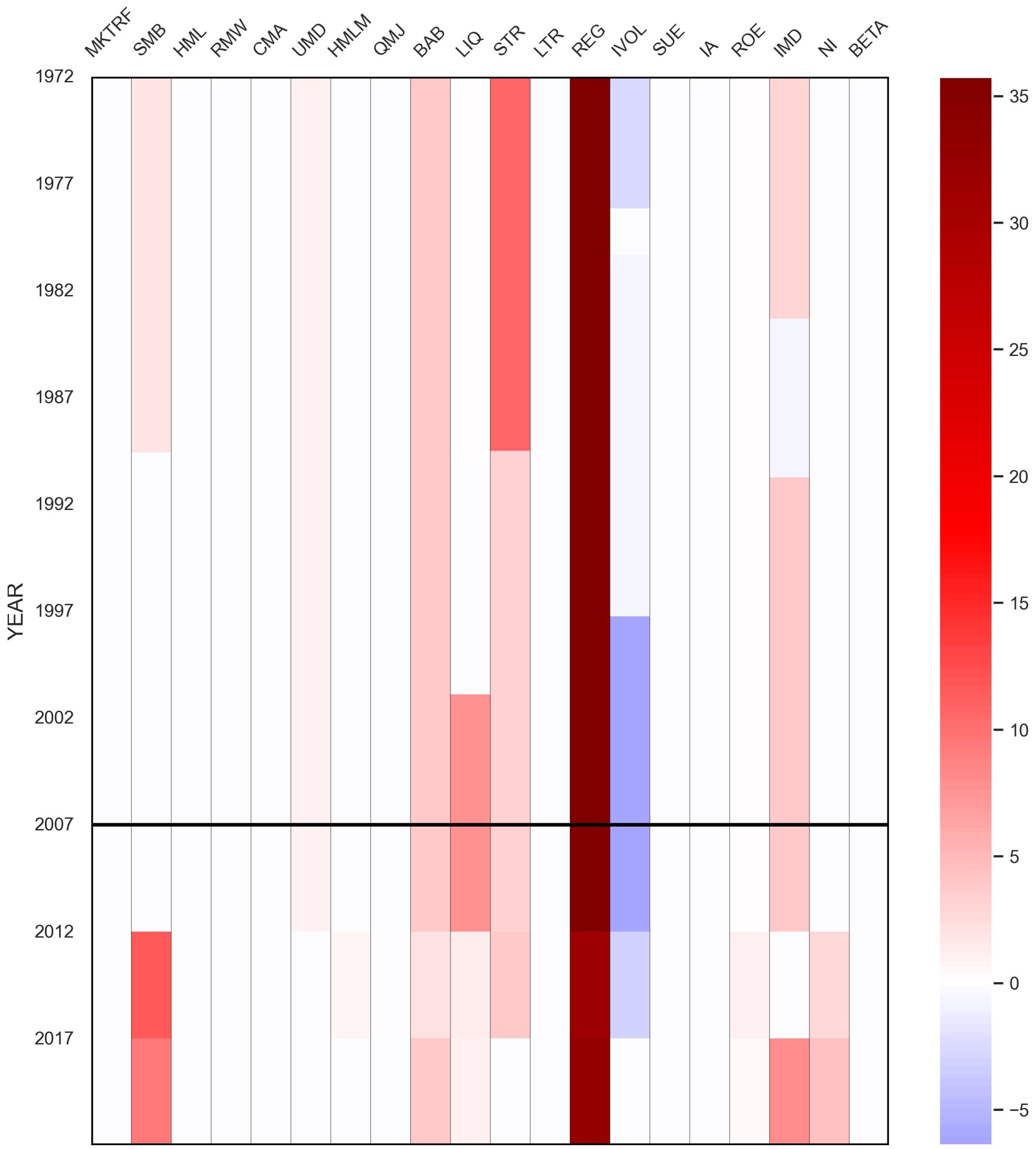


Figure 3: Evaluating and Estimating New Factors since Publication

Note: This figure displays a heatmap plotting the time-varying estimates for selected factor risk prices (SDF weights). A factor is only considered in the selection when discovered in academia (the sample end year in the published paper). The factor names and their sample end year are printed next to the year axis. In the color bar, the zero-value color of white indicates the factor is not selected for this period.

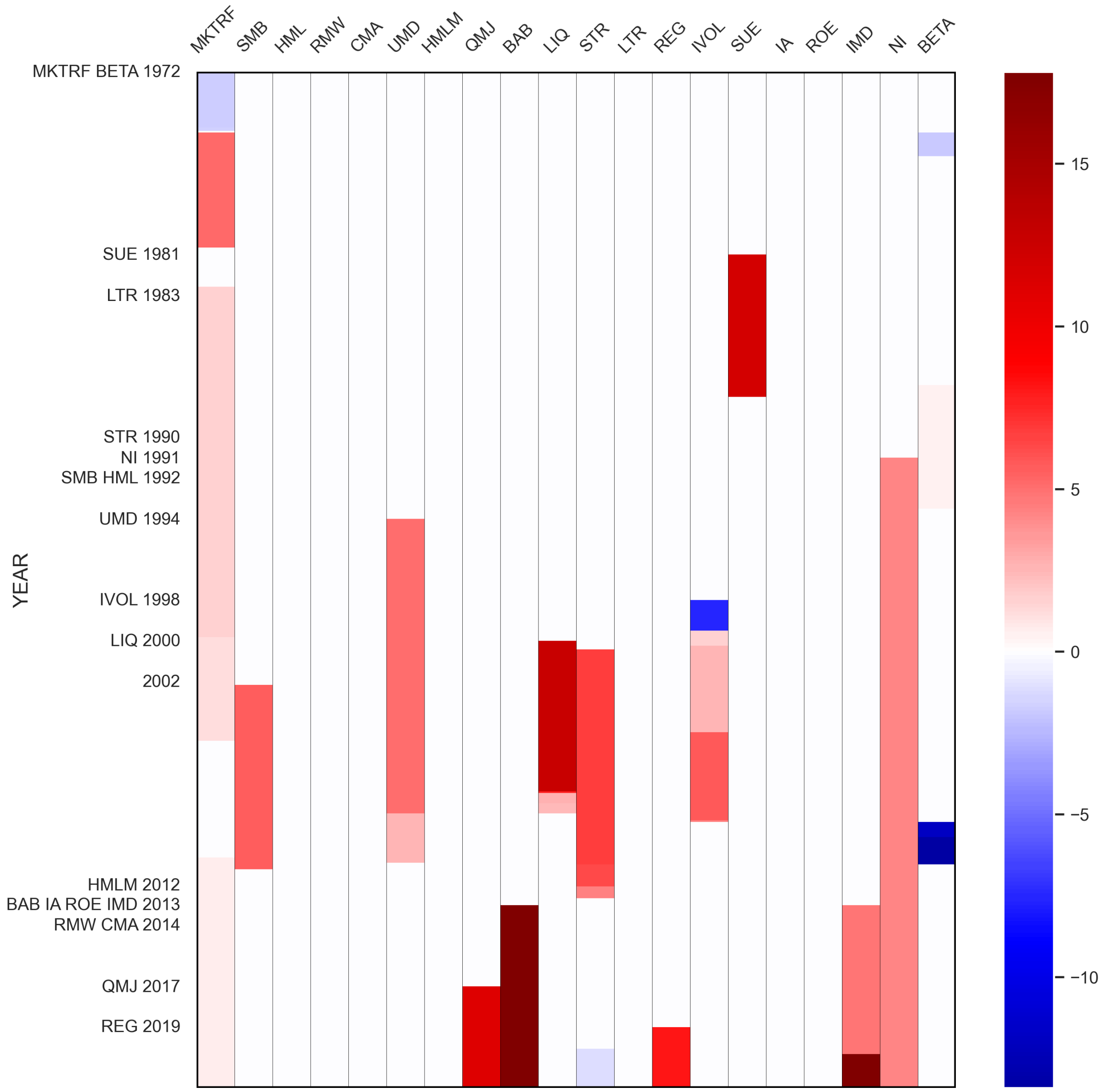


Table 1: Estimates for Selected Factor Risk Prices

Note: This table reports the identities of the selected factors, the number of single or multiple selection periods (yymm), and their risk price piecewise-constant estimates, summarizing the heatmap information in Figure 2. The training period is from 1972 to 2006, and the test period is from 2007 to 2021.

MKTRF				
SMB	1.70 (7201-8907)	11.67 (1201-1612)	9.32 (1701-2112)	
HML				
RMW				
CMA				
UMD	0.98 (7201-1112)			
HMLM	0.75 (1201-1612)			
QMJ				
BAB	3.74 (7201-1112)	2.19 (1201-1612)	3.70 (1701-2112)	
LIQ	0.24 (7201-9001)	7.76 (0012-1112)	1.37 (1201-1612)	1.14 (1701-2112)
STR	10.59 (7201-8906)	3.40 (8907-1112)	3.74 (1201-1612)	
LTR				
REG	35.76 (7201-1112)	31.72 (1201-1612)	33.08 (1701-2112)	
IVOL	-2.49 (7201-7802)	-0.59 (8005-9703)	-6.40 (9704-1112)	-3.12 (1201-1612)
SUE				
IA				
ROE	0.33 (7201-1112)	0.90 (1201-1612)	0.59 (1701-2112)	
IMD	3.14 (7201-8304)	-0.50 (8305-9009)	3.82 (9010-1112)	8.01 (1701-2112)
NI	2.57 (1201-1612)	4.34 (1701-2112)		
BETA				

Table 2: (Conditional) Asset Pricing Performance Comparison: HJ-R^2 (in %)

Notes: This table reports the asset pricing performance for the training sample from 1972 to 2006. As described in Section 4.1, we provide results for unconditional HJ-R^2 and conditional HJ-R^2, which is based on each different aggregate predictor described in Section 3.1. Our SFGMM is estimated using all these aggregate predictors. We also provide an all-factor model estimated with and without IV, and other time-varying coefficient models (introduced in the simulation study). Other ad-hoc selected models, including Fama-French three- and five-factor models, and the $q5$ q-factor model, are estimated with IV by GMM.

Models	Unconditional	tbl	infl	tms	dfy	ep	svar	ni	dy	lev	liq
					Time-Varying Coefficient Model						
SFGMM	87	93	89	89	87	91	85	76	93	89	72
AGMM	86	92	91	90	90	90	86	88	96	93	81
LGMM	78	92	86	62	84	79	43	90	82	67	81
					Constant-Coefficient Model by GMM						
FF3	50	56	32	53	68	71	57	30	72	57	20
FF5	55	69	56	56	71	71	60	20	72	49	41
q5	64	83	69	58	78	76	69	62	80	67	7
all factors	79	82	85	81	91	91	84	87	93	90	70
all factors (without IV)	91	76	83	92	-15	77	-58	82	89	86	-25

Table 3: Risk Factor Importance across Macroeconomic Regimes

Notes: We present the top five factors chosen by our time-varying specified factor model and their average risk price estimates under different macroeconomic regimes (high or low for each macro variable). The macro variables, which are listed in the table, comprise the 3-month treasury bill rate, inflation, TERM and default factor yields, as well as various aggregate equity market characteristics such as earnings-to-price ratio, stock variance, net equity issues, dividend yields, leverage, and liquidity.

	High State					Low State				
tbl	REG (34.84)	STR (6.38)	IMD (4.79)	SMB (4.14)	BAB (3.72)	REG (34.79)	STR (4.49)	IVOL (-3.58)	LIQ (3.53)	BAB (3.39)
infl	REG (35.45)	STR (6.29)	IMD (4.00)	BAB (3.74)	IVOL (-3.07)	REG (35.12)	STR (5.63)	BAB (3.52)	IVOL (-2.93)	SMB (2.41)
tms	REG (35.76)	STR (7.58)	BAB (3.74)	IVOL (-2.92)	IMD (2.68)	REG (34.84)	IMD (4.45)	STR (4.26)	BAB (3.67)	SMB (3.56)
dfy	REG (35.39)	STR (6.76)	BAB (3.63)	IVOL (-3.61)	IMD (3.34)	REG (34.95)	STR (3.81)	BAB (3.59)	IMD (3.32)	SMB (2.83)
ep	REG (35.39)	STR (7.65)	BAB (3.60)	IVOL (-3.27)	IMD (3.07)	REG (35.16)	STR (4.90)	BAB (3.64)	IMD (2.92)	IVOL (-2.74)
svar	REG (35.54)	STR (5.62)	IVOL (-3.97)	BAB (3.74)	IMD (3.51)	REG (35.34)	STR (4.71)	BAB (3.70)	IMD (3.44)	IVOL (-2.11)
ni	REG (35.42)	STR (5.29)	BAB (3.71)	IMD (3.64)	IVOL (-2.98)	REG (34.83)	STR (5.17)	LIQ (3.75)	IVOL (-3.68)	BAB (3.45)
dy	REG (35.72)	STR (7.70)	BAB (3.74)	IMD (3.42)	IVOL (-3.21)	REG (35.28)	STR (4.28)	BAB (3.71)	IMD (3.57)	IVOL (-2.02)
lev	REG (35.76)	STR (8.49)	BAB (3.74)	IMD (3.18)	IVOL (-2.72)	REG (34.29)	SMB (4.82)	IMD (3.90)	BAB (3.45)	STR (3.16)
liq	REG (34.81)	STR (5.69)	SMB (3.55)	BAB (3.49)	IMD (2.53)	REG (35.43)	STR (5.15)	IVOL (-4.21)	IMD (3.77)	BAB (3.74)

Table 4: Dissecting Anomalies across Macroeconomic Regimes

Notes: We present the unconditional HJ-R^2 (8) of our time-varying specified factor model across different macroeconomic regimes (high or low for each macro variable). The macro variables, including the 3-month treasury bill rate, inflation, TERM and default factor yields, and various aggregate equity market characteristics (earnings-to-price ratio, stock variance, net equity issues, dividend yields, leverage, and liquidity), are listed in the table. The size and anomaly bivariate-sorted portfolios comprise book-to-market ratio, operating profitability, investment, momentum, beta, short-term reversal, long-term reversal, return variance, residual variance, net issuance, and accruals. The number of monthly observations is provided for each high and low state.

Portfolios	tbl	infl	tms	dfy	ep	svar	ni	dy	lev	liq
					High State					
me-bm 5×5	92	51	80	72	85	87	21	25	54	-73
me-op 5×5	93	73	61	71	84	76	-15	23	50	33
me-inv 5×5	93	71	77	78	85	90	38	28	57	-5
me-mom 5×5	96	81	79	75	88	71	76	23	48	27
me-beta 5×5	92	19	64	75	67	75	-9	17	39	-11
me-str 5×5	94	83	67	78	88	82	44	27	48	-2
me-ltr 5×5	85	25	76	84	89	76	59	33	60	-17
me-var 5×5	93	69	51	79	80	89	57	24	29	49
me-revar 5×5	92	72	52	78	79	87	58	24	31	54
me-ni 5×7	95	81	73	80	82	82	11	23	47	20
me-ac 5×5	97	55	3	72	78	78	12	13	41	26
#obs	58	59	78	59	82	90	94	57	83	70
					Low State					
me-bm 5×5	88	-106	42	63	78	83	62	86	89	39
me-op 5×5	82	-161	42	69	74	83	70	84	85	36
me-inv 5×5	89	-123	43	71	77	78	68	81	88	52
me-mom 5×5	86	-54	59	64	57	75	71	62	62	58
me-beta 5×5	89	-75	42	73	74	71	69	70	91	31
me-str 5×5	83	-106	65	65	63	68	74	74	87	69
me-ltr 5×5	87	-309	34	80	54	75	71	76	70	0
me-var 5×5	81	11	57	34	68	66	54	85	92	49
me-revar 5×5	83	12	58	36	69	69	59	84	91	50
me-ni 5×7	84	-85	50	59	77	56	65	78	86	54
me-ac 5×5	79	-220	50	57	78	79	73	85	92	58
#obs	71	68	57	81	107	51	67	136	82	69

Table 5: Risk-Adjusted Investment Performance Comparison

Notes: Using the estimated SDF models, this table reports the risk-adjusted investment performance (ASR for annualized Sharpe ratio, Alpha for monthly Jensen's alpha, and MDD for one-year maximum drawdown). We estimate the SDF model from 1972 to 2006 and perform an out-of-sample evaluation from 2007 to 2021. Results for restricted (long-only) and unrestricted SDF-implied factor investments are provided. We also provide an all-factor model estimated with and without IV, and other time-varying coefficient models (introduced in the simulation study). Other ad-hoc selected models, including Fama-French three- and five-factor models, and the $q5$ q-factor model, are estimated with IV by GMM. Finally, results for a buy-and-hold excess market return, equally-weighted, and mean-variance efficient strategies using all factors are also provided.

	Restricted (Long-Only)			Unrestricted		
	ASR	Alpha	MDD	ASR	Alpha	MDD
Panel A: In-Sample (1972-2006)						
Time-Varying Coefficient Model						
SFGMM	2.59	1.00***	6.15	2.95	0.98***	3.16
AGMM	1.92	0.76***	9.76	2.86	0.98***	2.52
LGMM	1.64	0.75***	12.52	3.09	1.06***	2.81
Constant-Coefficient Model by GMM						
FF3	0.74	0.25***	16.37	0.74	0.25***	16.37
FF5	1.20	0.35***	6.09	1.11	0.31***	7.43
$q5$	2.11	0.63***	5.15	2.03	0.62***	5.52
all factors	1.67	0.74***	15.75	2.63	1.09***	4.74
all factors (without IV)	1.97	0.68***	10.18	2.91	0.84***	3.30
Other Strategies						
MKTRF	0.39	-	45.31	0.39	-	45.31
EW	1.37	0.64***	14.36	1.37	0.64***	14.36
MVE	2.13	0.67***	10.07	3.23	0.81***	2.21
Panel B: Out-of-Sample (2007-2021)						
Time-Varying Coefficient Model						
SFGMM	0.97	0.31***	5.53	0.86	0.25***	9.89
AGMM	0.58	0.15**	7.86	0.29	0.16	19.77
LGMM	0.51	0.13*	5.61	0.36	0.07	23.92
Constant-Coefficient Model by GMM						
FF3	0.38	−0.18*	27.06	0.38	−0.18*	27.06
FF5	0.64	0.56	9.27	0.71	0.11	8.35
$q5$	1.26	0.25***	5.84	1.18	0.23***	7.08
all factors	0.89	0.32***	8.14	1.11	0.59**	12.73
all factors (without IV)	1.00	0.21***	4.67	1.22	0.33***	4.67
Other Strategies						
MKTRF	0.68	-	46.34	0.68	-	46.34
EW	0.47	0.64***	9.89	0.47	0.64***	9.89
MVE	1.15	0.67***	3.87	1.36	0.81***	4.59

Appendices

The appendix is organized as follows: Section A presents simulation studies validating the applicability of SFGMM. Section B provides the primitive conditions for the paper's main results and theoretical guarantees on the global property.

A Simulation Study

We conduct simulation studies to evaluate the finite sample performance of SFGMM in a panel regression setting:

$$y_{t,i} = X'_{t,i}\gamma_{o,t} + u_{t,i} \quad t = 1, 2, 3, \cdots, T;\ i = 1, 2, 3, \cdots, K, \tag{A-1}$$

where $\gamma_{o,t} = (\gamma_{o,t,1}, \cdots, \gamma_{o,t,p})'$ and $X_{t,i}$ is a $p \times 1$ vector generated from multivariate normal distribution $N(\mathbf{0}, \Sigma)$ with $\Sigma_{j,k} = 0.3$ for $j \neq k$ and $\Sigma_{j,j} = 1$ for $j, k = 1, \cdots, p$. $u_{t,i} = \rho_1 \epsilon_{t-1,i} + \epsilon_{t,i}$ with $\epsilon_{t,i} \sim i.i.d.N(0, 1)$. $z_t \in R^l$ includes the conditioning variables orthogonal to u_t that satisfy $z_t = \rho_2 z_{t-1} + v_t$, where $z_0 = 0$, and $v_{t,i} \sim i.i.d.N(0, 1)$. We construct the IVs $\tilde{z}_t$ using orthogonal basis functions in the Hilbert space $L^2(0, 1)$. Let $\Psi(z) = \big(\psi_1(z), \cdots, \psi_m(z)\big)'$, where $\psi_j(z) = \sqrt{2}\cos(\pi j z)$ for $1 \leq j \leq m$ and $\tilde{z}_t = \big(1, \Psi'(z_{t,1}), \Psi'(z_{t-1,1}), \Psi'(z_{t-2,1}), \cdots, \Psi'(z_{t,l}), \Psi'(z_{t-1,l}), \Psi'(z_{t-2,l})\big)' \in R^{\tilde{l}}$ with $\tilde{l} = 3ml + 1$. We also examine the performance of SFGMM under varying degrees of serial dependency by setting $\rho_1 \in [0, 0.95]$ and $\rho_2 = 0.5$.

We examine the finite sample performance of SFGMM under two different sample sizes, $T = 120$ and $T = 240$, and two cross-sectional dimensions, $K = 20$ and $K = 50$, respectively. We further consider a 60-observation out-of-sample period. For each sample size T, we set $m = 3$ and $l = 11$, resulting in $\tilde{l} = 100$. In practice, we determine the tuning parameter values through repeated 5-fold CV, as detailed in the appendix. Consequently, the number of moment conditions utilized in estimation is $r = 4/5\tilde{r}$, where $\tilde{r} = 2000$ for $K = 20$ and $\tilde{r} = 5000$ for $K = 50$. Our investigation in various scenarios enhances our understanding of SFGMM's finite sample performance.

Data Generating Process (DGP).

$$\gamma_{o,t,i} = \sum_{j=1}^{3} a_{i,j} \mathbb{1}\left(\frac{(j-1)T}{3} + 1 \leq t \leq \frac{jT}{3}, i \in \Upsilon_j\right), \tag{A-2}$$

where each Υ_j contains 10 random draws from $\{1, 2, \cdots, p/2\}$ with $j = 1, 2, 3$ and $a_{i,j} \sim U[1, 5]$. $\mathbb{1}(\cdot)$ is the indicator function that takes one if the condition inside is satisfied and zero otherwise. To allow for time-varying model specifications, in each period, we randomly choose 10 active covariates with nonzero coefficient values and let at least $p/2$ covariates have zero coefficient values during the whole sample period. As a result, we have two different numbers of extra inactive covariates: 40 for $p = 50$ and 90 for $p = 100$.

In this study, we compare SFGMM with several leading alternative estimation methods: constant GMM (CGMM) (Hansen, 1982), local linear GMM (LGMM) (Lewbel, 2007), and afine GMM (AGMM) (Nagel and Singleton, 2011). LGMM and AGMM model time-varying parameter values as functions of scaled time, non-parametrically and parametrically, respectively. Performance evaluation involves out-of-sample comparisons with a scheme similar to Lettau and Pelger (2020). Specifically, we estimate parameters up to time t using the previous T observations, then obtain one-step-ahead forecasts for the parameter's value at $t + 1$.

We evaluate the accuracy of different estimation methods using the following performance metrics: average absolute estimation error (Abs.Err.) $(pT)^{-1} \sum_{t=1}^{T} \|\hat{\gamma}_t - \gamma_{o,t}\|$ and average squared moment errors (Mom.Err.) $(KT)^{-1} \sum_{t=1}^{T} \hat{u}_t' \hat{u}_t$, where $T = 120$ or 240 for in-sample studies and $T = 60$ for out-of-sample comparisons.

Table A.1 presents estimation comparisons for SFGMM and other methods under different scenarios. We report the average norms and standard deviations based on 50 replications. CGMM suffers from model misspecification due to the constant parameter assumption. Even though AGMM uses a linear function of scaled time to specify time-varying parameters, it still generates non-negligible misspecification issues. LGMM performs better than CGMM and AGMM but is still inferior to SFGMM. SFGMM, with its ability to handle time-varying sparsity and unknown structural breaks, provides a more accurate estimation and outperforms other methods in various norms.

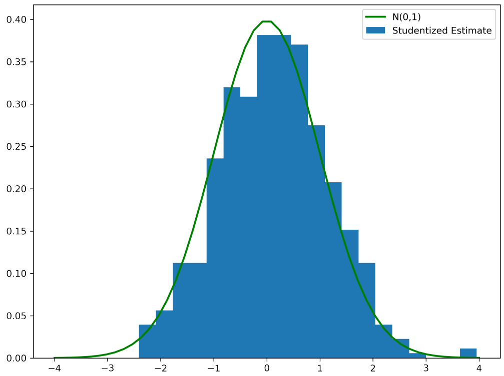
The low standard errors of SFGMM further support its reliability and efficiency.

Table A.2 confirms the robust out-of-sample performance of SFGMM, demonstrating its robustness and effectiveness across various moment conditions, sample sizes, and degrees of serial dependency. In addition, Table A.3 presents the results of our model selection analysis for SFGMM, utilizing two criteria: the accuracy rate (% correct) and the recall rate (relevant included), respectively. The accuracy rate measures the proportion of correctly selected estimates, while the recall rate measures the proportion of effective (nonzero) factors included in the selection. By achieving high accuracy and recall scores across different simulation settings, Table A.3 demonstrates our estimator's superior ability to consistently estimate time-varying parameter values, detect structural changes, and select time-varying sets of effective factors while avoiding useless ones. Specifically, the results are consistently better as T and N increase, which aligns with our theoretical analysis.

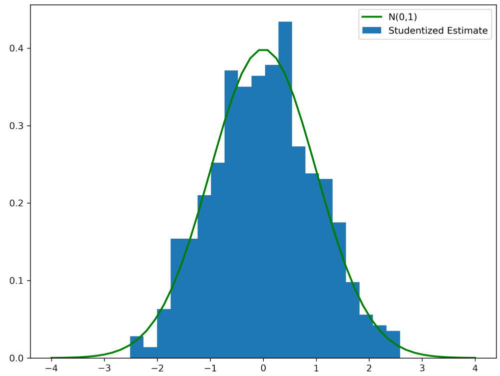
Finally, Figure A.1 displays histograms of the studentized time-varying estimates for $(\rho_1, \rho_2) = (0, 0.5)$. According to asymptotic theory, the studentized estimates should follow an $N(0, 1)$ distribution. Despite serial dependency, the histograms provide an excellent approximation to the $N(0, 1)$ distribution in finite samples.

Figure A.1: Simulation: Asymptotic Normality

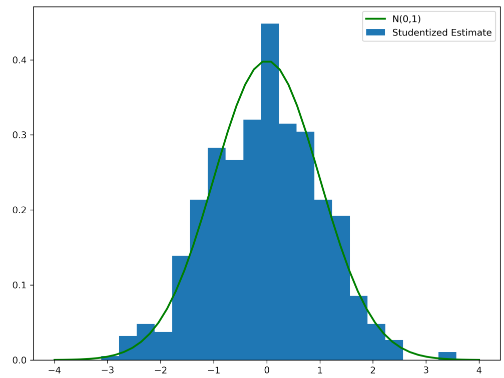
Note: This figure displays histograms of the studentized time-varying estimates for the first nonzero $\hat{\gamma}_{t,i}$ from bootstrap samples. The figure confirms that the studentized estimates approximate an $N(0,1)$ distribution even with serial dependency.



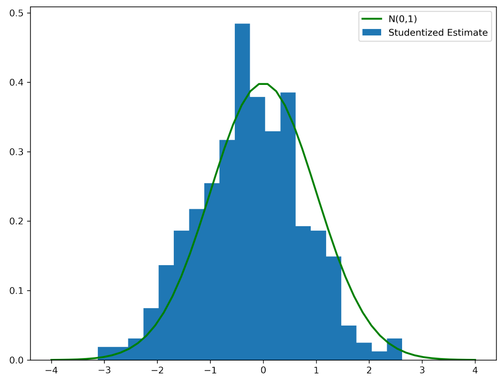
(a) T=120, K=20



(b) T=120, K=50



(c) T=240, K=20



(d) T=240, K=50

Table A.1: Simulation: In-Sample Estimation Accuracy

Note: In Section A, we compare the in-sample performance of SFGMM with CGMM, LGMM, and AGMM for various parameters. Specifically, we vary the length of periods T (120 and 240), number of regressors p (50 and 100), autocorrelation level ρ_1 (0 and 0.95), and number of cross-sectional observations K (20 and 50). We report the averages and standard deviation in parentheses of two different norms of fitting errors from 50 replications: average estimation error and average moment error.

	Criteria	SFGMM	CGMM	LGMM	AGMM	SFGMM	CGMM	LGMM	AGMM
		K=20				K=50			
$\rho_1 = 0$ T=120 p=50	Abs.Err	0.113	1.033	0.332	0.827	0.068	1.029	0.308	0.807
		(0.008)	(0.005)	(0.010)	(0.008)	(0.003)	(0.007)	(0.005)	(0.006)
	Mom.Err	1.159	50.869	4.428	31.889	0.980	52.307	5.063	32.349
		(0.046)	(3.375)	(0.427)	(2.592)	(0.020)	(2.443)	(0.332)	(2.072)
$\rho_1 = 0$ T=120 p=100	Abs.Err	0.128	1.147	0.351	0.913	0.082	1.024	0.310	0.877
		(0.009)	(0.005)	(0.011)	(0.008)	(0.003)	(0.008)	(0.006)	(0.006)
	Mom.Err	1.359	49.779	4.228	30.259	1.213	48.961	4.863	31.249
		(0.049)	(3.381)	(0.435)	(2.499)	(0.021)	(2.512)	(0.324)	(2.062)
$\rho_1 = 0$ T=240 p=50	Abs.Err	0.135	1.025	0.290	0.814	0.077	1.019	0.284	0.797
		(0.008)	(0.005)	(0.006)	(0.008)	(0.006)	(0.005)	(0.004)	(0.004)
	Mom.Err	1.293	52.748	4.426	33.113	1.189	52.073	4.657	32.097
		(0.071)	(3.709)	(0.378)	(3.128)	(0.038)	(1.991)	(0.232)	(1.662)
$\rho_1 = 0$ T=240 p=100	Abs.Err	0.132	1.127	0.348	0.912	0.081	1.124	0.351	0.862
		(0.008)	(0.005)	(0.007)	(0.006)	(0.003)	(0.007)	(0.006)	(0.005)
	Mom.Err	1.442	47.137	4.180	29.757	1.301	47.859	4.638	31.092
		(0.051)	(3.159)	(0.427)	(2.396)	(0.020)	(2.407)	(0.294)	(1.983)
$\rho_1 = 0.95$ T=120 p=50	Abs.Err	0.115	1.092	0.335	0.831	0.074	1.051	0.323	0.816
		(0.009)	(0.005)	(0.010)	(0.008)	(0.004)	(0.007)	(0.005)	(0.006)
	Mom.Err	1.259	51.293	4.506	31.977	0.991	52.474	5.129	33.892
		(0.049)	(3.383)	(0.427)	(2.700)	(0.023)	(2.536)	(0.401)	(2.174)
$\rho_1 = 0.95$ T=120 p=100	Abs.Err	0.131	1.157	0.358	0.920	0.085	1.025	0.311	0.876
		(0.010)	(0.006)	(0.012)	(0.008)	(0.004)	(0.008)	(0.007)	(0.007)
	Mom.Err	1.377	49.892	4.308	31.091	1.302	49.215	4.873	31.595
		(0.052)	(3.403)	(0.443)	(2.530)	(0.024)	(2.609)	(0.344)	(2.123)
$\rho_1 = 0.95$ T=240 p=50	Abs.Err	0.137	1.030	0.305	0.872	0.081	1.092	0.297	0.813
		(0.009)	(0.006)	(0.007)	(0.008)	(0.006)	(0.005)	(0.005)	(0.006)
	Mom.Err	1.299	53.281	4.516	33.717	1.203	52.473	5.666	32.395
		(0.073)	(3.759)	(0.398)	(3.411)	(0.043)	(1.993)	(0.240)	(1.674)
$\rho_1 = 0.95$ T=240 p=100	Abs.Err	0.135	1.140	0.358	0.937	0.084	1.183	0.371	0.880
		(0.009)	(0.006)	(0.007)	(0.006)	(0.004)	(0.007)	(0.007)	(0.005)
	Mom.Err	1.478	48.337	4.533	30.151	1.351	48.295	4.889	31.724
		(0.055)	(3.203)	(0.312)	(2.256)	(0.023)	(2.571)	(0.308)	(1.995)

Table A.2: Simulation: Out-of-Sample Estimation Accuracy

Note: For the DGP in Section A, we conduct a simulation study to compare the out-of-sample finite sample performance of SFGMM with other methods, including CGMM, LGMM, and AGMM. The study varies the length of periods T (120 and 240), the number of regressors p (50 and 100), autocorrelation level ρ_1 (0 and 0.95), and the number of cross-sectional observations K (20 and 50). Based on 60 time periods beyond T, we report the averages and standard deviation in parentheses of two different norms of fitting errors from 50 replications: average estimation and average moment errors.

	Criteria	SFGMM	CGMM	LGMM	AGMM	SFGMM	CGMM	LGMM	AGMM
			K=20				K=50		
$\rho_1 = 0$ T=120 p=50	Abs.Err	0.034	1.023	0.433	0.935	0.025	1.001	0.302	0.905
		(0.002)	(0.004)	(0.011)	(0.009)	(0.003)	(0.007)	(0.005)	(0.008)
	Mom.Err	0.962	46.991	4.013	30.849	0.873	51.283	4.997	31.462
		(0.047)	(3.484)	(0.459)	(2.531)	(0.029)	(2.267)	(0.329)	(2.168)
$\rho_1 = 0$ T=120 p=100	Abs.Err	0.102	1.152	0.329	0.890	0.046	0.984	0.293	0.722
		(0.008)	(0.007)	(0.013)	(0.008)	(0.004)	(0.007)	(0.006)	(0.007)
	Mom.Err	1.263	47.806	4.317	29.192	1.012	47.861	4.904	31.252
		(0.051)	(3.419)	(0.368)	(2.572)	(0.035)	(2.300)	(0.346)	(2.179)
$\rho_1 = 0$ T=240 p=50	Abs.Err	0.092	1.015	0.293	0.889	0.055	1.092	0.232	0.713
		(0.009)	(0.005)	(0.005)	(0.008)	(0.008)	(0.005)	(0.004)	(0.006)
	Mom.Err	1.192	51.498	3.456	32.139	1.042	52.002	3.983	31.074
		(0.068)	(3.814)	(0.422)	(3.593)	(0.041)	(2.560)	(0.397)	(2.474)
$\rho_1 = 0$ T=240 p=100	Abs.Err	0.135	1.017	0.329	0.888	0.086	1.117	0.343	0.801
		(0.011)	(0.009)	(0.06)	(0.006)	(0.005)	(0.007)	(0.008)	(0.005)
	Mom.Err	1.342	47.230	4.596	29.322	1.235	47.012	4.333	30.985
		(0.073)	(3.444)	(0.610)	(2.933)	(0.032)	(2.407)	(0.294)	(2.101)
$\rho_1 = 0.95$ T=120 p=50	Abs.Err	0.036	1.026	0.430	0.954	0.027	1.001	0.312	0.940
		(0.003)	(0.004)	(0.011)	(0.008)	(0.006)	(0.007)	(0.006)	(0.008)
	Mom.Err	0.993	46.972	4.019	30.239	0.904	51.293	4.998	31.498
		(0.053)	(3.490)	(0.558)	(2.622)	(0.037)	(2.188)	(0.352)	(2.206)
$\rho_1 = 0.95$ T=120 p=100	Abs.Err	0.115	1.172	0.383	0.810	0.072	0.990	0.298	0.722
		(0.010)	(0.006)	(0.012)	(0.008)	(0.004)	(0.008)	(0.007)	(0.007)
	Mom.Err	1.292	47.964	4.375	29.205	1.003	49.745	4.873	30.992
		(0.053)	(3.582)	(0.343)	(2.661)	(0.038)	(2.362)	(0.344)	(2.291)
$\rho_1 = 0.95$ T=240 p=50	Abs.Err	0.098	1.072	0.305	0.891	0.058	1.091	0.222	0.751
		(0.010)	(0.005)	(0.006)	(0.008)	(0.009)	(0.008)	(0.004)	(0.006)
	Mom.Err	1.231	51.368	3.574	32.420	1.086	52.732	3.946	31.153
		(0.072)	(3.925)	(0.447)	(3.598)	(0.050)	(2.569)	(0.402)	(2.407)
$\rho_1 = 0.95$ T=240 p=100	Abs.Err	0.139	1.025	0.332	0.897	0.080	1.117	0.343	0.819
		(0.012)	(0.009)	(0.005)	(0.006)	(0.005)	(0.007)	(0.007)	(0.006)
	Mom.Err	1.359	47.380	4.643	29.713	1.243	47.012	4.423	28.231
		(0.077)	(3.251)	(0.619)	(2.853)	(0.031)	(2.420)	(0.302)	(2.119)

Table A.3: Simulation: Selection Consistency

Note: This table reports the selection rates for SFGMM, measured by two criteria: % correct (accuracy) and Relevant included (recall). % correct measures the proportion of correctly selected estimates, while Relevant included measures the proportion of truly (nonzero) factors included. We report these measures' average and standard deviation in parentheses based on 50 iterations in the in-sample period.

	$(\rho_1, \rho_2) = (0, 0.5)$			
	T=120		T=240	
	% correct	Relevant included	% correct	Relevant included
40 extra factors, K=20	0.87	1.00	0.92	1.00
	(0.02)	(0.00)	(0.01)	(0.00)
40 extra factors, K=50	0.97	1.00	0.97	1.00
	(0.01)	(0.00)	(0.01)	(0.00)
90 extra factors, K=20	0.94	1.00	0.95	1.00
	(0.02)	(0.00)	(0.01)	(0.00)
90 extra factors, K=50	0.98	1.00	0.99	1.00
	(0.01)	(0.00)	(0.00)	(0.00)
	$(\rho_1, \rho_2) = (0.95, 0.5)$			
40 extra factors, K=20	0.85	1.00	0.92	1.00
	(0.02)	(0.00)	(0.01)	(0.00)
40 extra factors, K=50	0.96	1.00	0.96	1.00
	(0.02)	(0.00)	(0.02)	(0.00)
90 extra factors, K=20	0.94	1.00	0.95	1.00
	(0.02)	(0.00)	(0.01)	(0.00)
90 extra factors, K=50	0.96	1.00	0.98	1.00
	(0.01)	(0.00)	(0.01)	(0.00)

B Assumptions and Global Properties

In this section, we first provide the primitive conditions for the paper's main results in Theorems 1 - 5. Then, we establish the global properties of our SFGMM estimators with and without information on structural breaks in Theorems A.1 and A.2, respectively.

We first introduce some additional notations. For a vector a of dimension K and a $K \times K$ symmetric matrix A, denote $\|a\|_\infty = \max_{1\le j\le K} |a_j|$ and $\|A\|_2 = \max_{a\in R^K, \|a\|=1} \|Aa\|$. Let $\rho_{\min}(A)$, $\rho_i(A)$ and $\rho_{\max}(A)$ denote the minimum, the ith, and the maximum eigenvalue of a matrix A. For a $K_1 \times K_2$ matrix B, denote $\|B\|_{2,\infty} = \max_{1\le k\le K_1} \|B_{k,}\|$ with $\|B_{k,}\| = \sqrt{\sum_{1\le j\le K_2} B_{kj}^2}$. For the penalty functions $P_i(\cdot)$ with $i = \lambda$ and η, they satisfy $P_i(0) = 0$; $P_i'(\cdot)$ and $P_i''(\cdot)$ denote the first and second derivatives for penalty functions. Let $\mathbb{M}_G = \{\Gamma = (\gamma_1, \cdots, \gamma_T)' \in R^{T\times p} : \gamma_{t,j} = \gamma_{\tau,j}, \text{for}\, t,\tau \in G_n^j, j \in \{1,\cdots,p\}, n \in \{1,\cdots,N_j\}\}$. We record the minimum and maximum duration of regimes as $|G_{\min}| = \min_{j,n} |G_n^j|$ and $|G_{\max}| = \max_{j,n} |G_n^j|$ for $1 \le n \le N_j$ and $1 \le j \le p$. We can partition the common value parameter vector into two subvectors $\theta = (\theta_{\mathbb{A}}', \theta_{\mathbb{Z}}')'$, where $\theta_{\mathbb{A}} \in R^q$ and $\theta_{\mathbb{Z}} \in R^{N-q}$ contain the nonzero and zero elements of θ, respectively. We can partition the associated regime duration vector into two subvectors $|\mathbb{G}| = \{|G_1|, \cdots, |G_N|\} = (|\mathbb{G}_{\mathbb{A}}|', |\mathbb{G}_{\mathbb{Z}}|')'$. We consider the signal strength $d_T = \min_{1\le n\le q} |\theta_{o\mathbb{A},n}|/2$ as half of the minimum signal.

We denote the loss function under the prior information on structural breaks as $\bar{L}(\theta) = r^{-1}\|g_T(\theta)\|^2 + \sum_{n=1}^N |G_n| * P_\eta(|\theta_n|)$. We denote the loss function without the prior information on the timing of structural breaks as $L(\Gamma) = r^{-1}\|g_T(\Gamma)\|^2 + \sum_{j=1}^p \sum_{t=2}^T P_\lambda(|\gamma_{t,j} - \gamma_{t-1,j}|) + \sum_{j=1}^p \sum_{t=1}^T P_\eta(|\gamma_{t,j}|)$. To save notations, we use $e(U_t, \gamma_t) = e(U_t, \theta)$ exchangeably in the proof.

Assumption A.1. *The penalty function $P_\lambda(x)$ is symmetric, nondecreasing and concave in $x \in [0, \infty)$, and $P_\lambda(0) = 0$. For some $C > 0$, $P_\lambda(x)$ is a constant for $x > C\lambda$. $P_\lambda'(x)$ is increasing in $\lambda \in [0, \infty)$ and $P_\lambda'(x)$ is decreasing in $x \in [0, \infty)$ with $\lambda^{-1}P_\lambda'(0+) = 1$. The same requirements hold for $P_\eta(x)$.*

Assumption A.2. *(i) For all* $1 \le l \le r$*,* $\{U_t', e_l(U_t, \gamma_{o,t})\}'$ *is a stationary* α*-mixing process with mixing coefficients* $\{\alpha(j)\}$ *satisfying* $\sum_{j=1}^{\infty} \alpha(j)^{\delta/(2+\delta)} < c_1$ *for some* $0 < \delta < 1$ *and positive constant* c_1*. (ii) Some positive constants* c_2 *and* c_3 *exist such that* $E|e_l(U_t, \theta_o)|^{2+\delta} < c_2$ *and* $E|\partial e_l(U_t, \theta_o)/\partial \theta_i|^{2+\delta} < c_3$ *for all* $1 \le i \le N$*,* $1 \le l \le r$*, and* $1 \le t \le T$*. (iii) Positive constants* c_4 *and* ν_1 *exist such that for any* $\epsilon > 0$*,* $P\big(|e_l(U_t, \gamma_{o,t})| > \epsilon\big) \le 2\exp\big(-(\epsilon/c_4)^{\nu_1}\big)$ *for all* $1 \le l \le r$ *and* $1 \le t \le T$*. (iv) Positive constants* c_5 *and* ν_2 *exist such that* $\alpha(j) \le \exp(-c_5 j^{\nu_2})$ *for all* $1 \le j \le \infty$*.*

Assumption A.3. *(i) A unique parameter matrix* Γ_o *exists such that* $E\big[e(U_t, \gamma_{o,t})\big] = \mathbf{0}$ *for all* $1 \le t \le T$*. (ii)* $r \ge q$ *and* $Nrexp(-qT) = o(1)$*. (iii) For* $\bar{\nu}$ *that satisfies* $1/\nu_1 + 1/\nu_2 = 1/\bar{\nu} > 1$*,* $NrT\exp\big(-(qT)^{\bar{\nu}/2}\big) = o(1)$*.*

Assumption A.4. *(i) Measurable positive functions* $D_1(U_t)$ *and* $D_2(U_t)$ *with* $E\big[D_1^2(U_t)\big] < \infty$ *and* $E\big[D_2^2(U_t)\big] < \infty$ *exist such that for* $\theta_1, \theta_2 \in R^N$*, we have* $r^{-1/2}\|e(U_t, \theta_1) - e(U_t, \theta_2)\| \le D_1(U_t)\|\theta_1 - \theta_2\|$ *and* $r^{-1/2}\|\partial e(U_t, \theta_1)/\partial\theta - \partial e(U_t, \theta_2)/\partial\theta\| \le D_2(U_t)\|\theta_1 - \theta_2\|$ *for all* $1 \le t \le T$*. (ii) Positive constants* c_6 *and* c_7 *exist such that* $\rho_{\min}\Big\{\big[E\partial e_l(U_t, \theta_o)/\partial\theta_{\mathbb{A}}\big]\big[E\partial e_l(U_t, \theta_o)/\partial\theta_{\mathbb{A}}\big]'\Big\} \ge c_6$ *and* $\rho_{\max}\Big\{\big[E\partial e_l(U_t, \theta_o)/\partial\theta_{\mathbb{A}}\big]\big[E\partial e_l(U_t, \theta_o)/\partial\theta_{\mathbb{A}}\big]'\Big\} \le c_7$ *for all* $1 \le l \le r$ *and* $1 \le t \le T$*.*

Assumption A.5. *Suppose (i)* $P_\lambda'(d_T) = O\big(1/(\sqrt{T}|G_{\max}|)\big)$ *with* $d_T \ge \max\{\sqrt{q\log r/T}, \sqrt{q}\{\log[r(T+1)]\}^{1/\bar{\nu}}/T\} + \sqrt{q/T}$*. (ii)* $N = O(e^{T^\alpha})$ *for some* $\alpha \in (0, 1/2)$ *where* $N = \sum_{j=1}^{p} N_j$*. (iii) Let* $\mathbf{N}_0 = \{\theta_{\mathbb{A}} \in R^q : \|\theta_{\mathbb{A}} - \theta_{o\mathbb{A}}\| \le d_T\}$*, for* $\theta_{\mathbb{A}} \in \mathbf{N}_0$*,* $\max_n P''(\theta_{\mathbb{A},n}) = o\big(1/|G_{\max}|\big)$*. (iv)* $P_\lambda'(0+) \ge \sqrt{q/T}/|G_{\min}|$*. (v)* $\lambda \ge O\Big(\big[\eta/|G_n^j| + \sqrt{q\log r/T} - \sqrt{q}\{\log[r(T+1)]/T^{\bar{\nu}}\}^{1/\bar{\nu}} + \eta + \eta|G_{\min}|\big]/|G_{\min}|\Big)$*.*

Assumption A.6. *For any* $\epsilon > 0$*, a sufficiently small constant* $\nu_{T,\epsilon} > 0$ *with* $\nu_{T,\epsilon} = o(1/T)$ *that* $\inf_{\theta \in R^N, \|\theta - \theta_o\| > \epsilon} r^{-1}\|E\big[g_T(\theta)\big]\|^2 > \nu_{T,\epsilon}$ *exists.*

Assumption A.2 (i) and (ii) impose a mild requirement on serial dependency and moment conditions, which has been widely used in time series econometric literature (Chen and Hong, 2012). Assumption A.2 (iii) provides exponential bounds for sub-exponential mixing rates when the variables are not necessarily bounded. Such an assumption on the tail probability is widely used in high-dimensional econometric and machine learning studies. It includes a broad class of distributions, including continuous random variables with compact support, normal distributions, exponential distributions, etc. For example, when observations and moment conditions are independently and identically distributed, Dong et al. (2021) employ such a thin tail requirement when conducting variable selection with endogeneity and global smoothing nonparametric estimation in GMM. In contrast, our paper allows for dependent observations. It thus imposes an additional requirement on the rate that serial dependency decays to zero in Assumption A.2 (iv). Hence, Assumptions A.2 (iii) and (iv) ensure a Berstein's type bound on the tail probabilities for the score function when proving consistency and establishing asymptotic distribution.

Assumption A.3 (i) is a global identification assumption widely used in the GMM literature. It ensures the existence and uniqueness of the true parameter matrix Γ_o that satisfies the moment conditions. Assumption A.3 (ii) posits an identification requirement for the nonzero elements in the unknown parameter matrix Γ. Assumption A.3 (ii) and (iii) further provide a necessary condition for estimation consistency of SFGMM. It regulates the relationship between the number of moment conditions r, the sample size T, the total number of common parameter values N, and the actual number of nonzero values q shall satisfy. Notably, the rate requirement in (ii) is to bound the tail probability for partial sums of independent observations. At the same time, the further restriction in (iii) arises when we allow for serial dependency and is the price we need to pay when conducting time-varying parameter estimation and variable selection for high-dimensional dependent moment conditions.

Assumption A.4 (i) and (ii) are a Lipschitz condition on moment conditions and their first-order derivatives, which are widely used in the GMM literature (Han and Phillips, 2006). It is essential for establishing uniform convergence because it relates to stochastic equicontinuity (Newey, 1991). Assumption A.4 also ensures the applicability of our result to time-varying nonlinear GMM models. Cui et al. (2022) impose these conditions on estimating time-varying parameter values in high-dimensional GMM. Assumption A.4 (ii) originates from the fact that we must assume that the Jacobian matrix of the moment conditions has full column rank in a neighborhood of the true unknown value of parameters, which has been pervasive since Hansen (1982).

Assumption A.5 (i) helps provide an upper bound for the score function of the SFGMM criterion and further regulates the minimal signal strength so that nonzero elements and structural breaks can be detected and consistently estimated. Assumption A.5 (ii) is essential in ensuring the Hessian matrix of the SFGMM criterion is asymptotically positive definite. Assumption A.5 (iv) and (v) arise as additional requirements when we recover time-varying sparsity from unknown time-varying structures and ensure a local minimizer of the SFGMM criterion converges in probability to the oracle estimator. Assumption A.5 could be interpreted as a requirement that structural breaks are not so frequent that we could release the concerns about overfitting.

Assumption A.6 is a generalized condition on global identification. We relate this identification condition with the scaled squared norm due to the diverging dimension of moment restrictions. This relaxed condition can also be seen in Cui et al. (2021), which employs a high-dimensional set of moment conditions and a ridge fusion penalty to estimate time-varying parameters in GMM. Note that Assumption A.6 includes $\nu_{T,\epsilon} = \nu > 0$ as a special but stronger case, which is widely used in the GMM literature (Dong et al., 2021).

Theorem A.1. *Suppose Assumptions A.1-A.6 hold and the conditions in Theorem 1 hold. Consider* $q^{3/2}\log r = o(T)$*,* $q^{3/2}\big[\log[r(T+1)]\big]^{2/\bar{\nu}} = o(T^{2\bar{\nu}})$ *for* $1/\nu_1 + 1/\nu_2 = 1/\bar{\nu} > 1$ *and* $\max_{1\le n\le q} P_\eta(|\theta_{o\mathbb{A},n}|) = o\big(1/(q|G_{\max}|)\big)$*. Then, the local minimizer* $\tilde{\theta}$ *that solves (5) satisfies that for any* $\kappa_1 > 0$*, a positive finite* $\omega_1 > 0$ *exists such that*

$$\lim_{T\to\infty} P\big[\bar{L}(\tilde{\theta}) + \omega_1 < \inf_{\|\theta-\theta_o\|>\kappa_1} \bar{L}(\theta)\big] = 1.$$

Theorem A.2. *Suppose Assumptions A.1-A.6 hold and the conditions in Theorem 4 hold. Consider* $q^{3/2}\log r = o(T)$*,* $q^{3/2}[\log[r(T+1)]]^{2/\bar{\nu}} = o(T^{2\bar{\nu}})$ *for* $1/\nu_1 + 1/\nu_2 = 1/\bar{\nu} > 1$*,* $\max_{1\le n\le q} P_\eta(|\theta_{o\mathbb{A},n}|) = \big(1/(q|G_{\max}|)\big)$ *and* $P_\lambda(b_T) = o(1/q)$*. Then, the local minimizer* $\hat{\Gamma}$ *that solves (6) satisfies that for any* $\kappa_2 > 0$*, a positive finite* $\omega_2 > 0$ *exists such that*

$$\lim_{T\to\infty} P\big[L(\hat{\Gamma}) + \omega_2 < \inf_{\|\Gamma-\Gamma_o\|>\kappa_2} L(\Gamma)\big] = 1.$$

References

Chen, B. and Y. Hong (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica 80*(3), 1157–1183.

Cui, L., G. Feng, and Y. Hong (2022). Regularized GMM for time-varying models with application to asset pricing. Technical report, City University of Hong Kong.

Cui, L., Y. Hong, and Y. Li (2021). Solving euler equations via two-stage nonparametric penalized splines. *Journal of Econometrics 222*(2), 1024–1056.

Dong, C., J. Gao, and O. Linton (2021). High dimensional semiparametric moment restriction models. *Journal of Econometrics*.

Han, C. and P. C. B. Phillips (2006). GMM with many moment conditions. *Econometrica 74*(1), 147–192.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 1029–1054.

Lettau, M. and M. Pelger (2020). Estimating latent asset-pricing factors. *Journal of Econometrics 218*(1), 1–31.

Lewbel, A. (2007). A local generalized method of moments estimator. *Economics Letters 94*(1), 124–128.

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

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica*, 1161–1167.