Breaks and Trends in Factor Premia*

Liyuan Cui Guanhao Feng Jianxin Ma Yinan Su

First version: Jun. 2025; This version: Dec. 2025

Abstract

We estimate a cross-sectional asset pricing model with structural breaks and regime-dependent momentum trends in factor premia. The method incorporates total-variation (TV) regularization into a characteristic-based factor model, nesting the panel predictive regression and the Fama-MacBeth procedure as extreme cases. It produces piecewise linear trend lines of varying frequencies, capturing either long-term or fine-grained factor momentum trends depending on the TV penalty, which is tuned in a machine learning framework. Large-sample theoretical properties of the estimators include consistency and asymptotic normality. Out-of-sample trend-following, factor-timing strategies yield superior performance compared to buy-and-hold or factor momentum strategies with fixed windows. Time-varying factor selection and premia estimates improve SDF construction compared to the static counterparts.

Keywords: cross-sectional predictability; factor premia; momentum trends; structural breaks; time-varying sparsity.

*We thank Dacheng Xiu, Raymond Kan, Markus Pelger, Bryan Kelly, Jia Li, Yan Liu, Stefan Nagel, Nikolai Roussanov, Allan Timmermann, the seminar and conference participants at the 4th Hong Kong Conference for Fintech, AI, and Big Data in Business, 2nd HKUST IAS-SBM Joint Workshop on Financial Econometrics, and 2025 Tsinghua Machine Learning Application in Financial Economics Conference, for their valuable comments. Cui (E-mail: liyuan.cui@cityu.edu.hk) and Feng (E-mail: gavin.feng@cityu.edu.hk) are at the City University of Hong Kong; Ma (E-mail: jianxin.ma@warwick.ac.uk) is at the University of Warwick; Su (E-mail: ys@jhu.edu) is at Johns Hopkins University.

1 Introduction

Risk-return modeling of the stock returns mainly relies on factor models. While significant progress has been made in using firm characteristics to estimate dynamic factor loadings, estimating dynamic factor premia remains challenging due to the limitations of time series information.[1] Changes in factor premia, if they exist, are inevitably obscured by factor return realizations, which are of greater magnitude and frequency. However, identifying structural breaks and trends in factor premia is crucial for constructing the Stochastic Discount Factor (SDF) and for factor-based investments. New advances in this area are critically needed.[2]

From an economic perspective, there are many reasons why factor premia might change, and in particular, change in a break and trend fashion. Prior studies show that factor premia can decay or vanish after the publication of a factor's construction, can be influenced by factor-level trading flows, and may undergo abrupt changes in response to structural shifts in the underlying economic conditions, such as the Global Financial Crisis and the COVID-19 pandemic.[3] The variety of underlying reasons suggests that breaks and trends can emerge in complex ways, and that modeling these economic mechanisms individually may be insufficient to capture the full picture. The full picture is likely a blend of short- and long-term (or high- and low-frequency) variations. Therefore, a data-driven approach can be well-suited to tease out the breaks and trends from the stock return data.[4] We propose a new method that accomplishes these goals and demonstrate its general usefulness in both asset pricing and invest-

[1]The discussion of firm characteristics and factor loadings is in prior work. Fama and French (1993) link expected returns to factor betas, and Daniel and Titman (1997) examine firm characteristics. Recent work that integrates firm characteristics includes Avramov (2004), Fan et al. (2016), Kelly et al. (2019), and Chen, Roussanov, and Wang (2025).

[2]Dangl and Halling (2012), Smith and Timmermann (2022), Patton and Weller (2022), and Chib et al. (2023) among others also contribute to this area. We discuss the differences in the approach further below.

[3]For example, Pástor and Stambaugh (2001) identify structural breaks in the equity premium; Daniel and Moskowitz (2016) document momentum crashes during market rebounds; McLean and Pontiff (2016) show anomaly premia decline post-publication.

[4]As an alternative to the "data-driven" approach, one could also study time-varying risk premium by modeling its association with observable economic variables, for example An et al. (2025) associate factor premia with trading-flow-driven quantity fluctuation; Gagliardini et al. (2016) model risk premia as a linear function of lagged instruments such as the default spread.

ment applications.

Our method is called Total-Variation Panel Regression (TV-PR). It is designed with two components that facilitate the identification of dynamic factor premia and distinguish it from other approaches for the same purpose. First, we leverage the vast amount of data in the *panel* of stock returns and characteristics to improve the estimation of factor premia, rather than solely using the time series of factor realizations. We rely on the aforementioned factor literature that connects firm characteristics to factor loadings in order to focus on estimating factor premia. Then, estimation focuses on factor premia (γ_t), which, when multiplied by the loadings (characteristics), predict individual stock return ($r_{i,t+1}$).[5] As a result, the estimation boils down to a return *predictive panel regression*, where time-varying predictive coefficients are the time-varying factor premia. TV-PR's reliance on cross-sectional information is analogous to the Fama and MacBeth (1973) procedure, where cross-sectional regressions are the first step in estimating (static) factor premia.

Second, we employ a machine learning technique known as total variation (TV) regularization to capture trends and breaks in factor premia.[6] In a nutshell, it imposes regularization on the *variation* between each pair of factor premia coefficients adjacent in time, $|\gamma_t - \gamma_{t+1}|$, also known as fused Lasso. Thereby, it encourages piecewise-constant estimates of factor premia over time. Specifically, when break signals are weak, the regularization "fuses" a sequence of γ's together, capturing a constant trend. Otherwise, it allows for abrupt changes in the γ sequence to capture breaks.

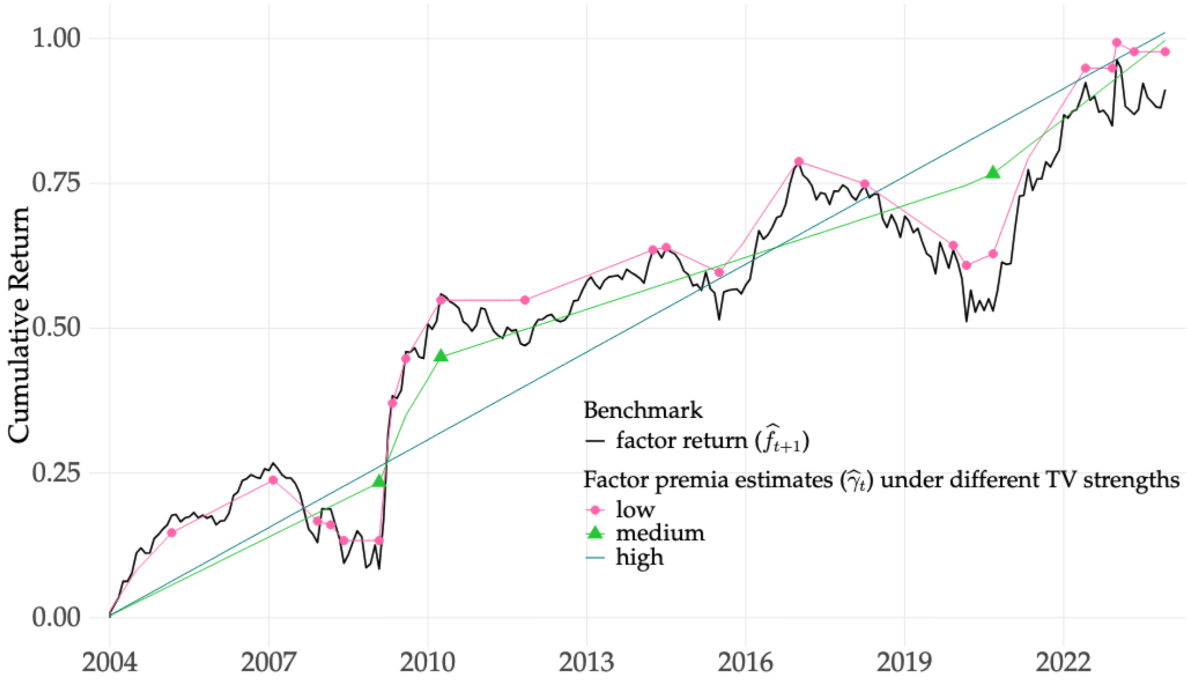
Figure 1 illustrates TV-PR's estimation of breaks and trends by plotting the cumulative sums of factor returns and factor premia estimates (where cumulative sums

[5]Specifically, we specify a "slope factor" model, where the characteristics serve as factor loadings. See Chib et al. (2023) for slope factor models. This setup is also related to the MSCI Barra model. We write γ_t in the Introduction for simplicity rather than $\gamma_{k,t}$, where k indexes factors. The main model is a multi-factor model, as detailed in Section 2.

[6]TV regularization is a widely used method in statistics and machine learning. Early contributions include Rudin et al. (1992) and Tibshirani et al. (2005). More recent developments include Harchaoui and Lévy-Leduc (2010), which introduces multiple change-point estimation; Li, Qian, and Su (2016), which introduces an adaptive fused-LASSO method to detect multiple structural breaks in panel data with interactive fixed effects; and Safikhani and Shojaie (2022), which applies TV regularization to jointly detect structural breaks and estimate high-dimensional VAR models. In asset pricing, Tang et al. (2022) apply a TV penalty to the cross-section in each period.

Figure 1: Breaks and trends in factor premia, different TV strengths/smoothness levels

Note: All curves are cumulative sums of (log) returns. Black: slope factor return ($\hat{f}_{t+1}$); Red/green/blue: factor premium estimates ($\widehat{\gamma}_t$) under low/medium/high TV regularization strength, respectively. In each constant γ segment, the cumulative sum is linear, indicating a trend. Higher TV regularization leads to fewer detected breaks and smoother estimated trends, and vice versa. This illustration displays the value factor (BE/ME); see other factors and multi-factor results further below. A detailed construction of every curve is presented in Section 4.



of piecewise constant γ estimates appear as piecewise linear segments). TV-PR effectively fits linear trend lines to match the overall direction of the factor performance. Drawing trend lines on factor performance charts is an intuitive idea, but often considered too simplistic due to the arbitrariness of choosing the change points. TV-PR formalizes this intuitive idea by selecting breaks with total-variation regularization in stock return predictive regressions.

TV-PR allows for the flexibility in capturing breaks of varying frequencies or, in other words, trends of varying scales. These choices are controlled by the TV regularization parameter. High regularization strength yields smoother trends (illustrated by the blue trend line in Figure 1), and, in the extreme, TV-PR reduces to fixed coefficients panel predictive regressions (e.g., Gu et al., 2020). On the other hand, with extremely low regularization, the model reduces to period-by-period cross-sectional regressions, and the γ_t estimate approaches the slope factor realization itself (the red jagged trend line approaching the black factor return curve). Thus, the TV regularization parameter opens up a spectrum of breaks and trends of varying frequencies/smoothness. Neither extreme, both of which are already studied in the literature and nested within our model, yields optimal performance in various investment and asset pricing applications, as detailed further below. Instead, we find "sweet spots" that lie between these two extremes, where the model admits just the right amount of breaks and trends to best suit the specific applications.[7] Such "sweet spots" are identified by tuning the TV regularization parameter using standard machine learning techniques.

TV-PR captures trends of varying smoothness adaptively, rather than using pre-specified trend lengths. As illustrated in Figure 1, the TV regularization parameter controls the overall scale of the trends, rather than directly fixing a uniform trend segment length. The lengths of the trend segments are adaptive to the data, subject to break estimation. This adaptability enables the detection of structural breaks more promptly, rather than relying on a fixed look-back window that may include data from a different regime. Fixed look-back windows are adopted in momentum trading strategies and in rolling-window re-estimates of the same model.[8] Empirical results reported further below support TV-PR's superior performance.

As such, TV-PR incorporates breaks and trends into an otherwise canonical framework, making it compatible with many existing empirical asset pricing techniques and applications. An important aspect we consider is the selection of factors. It is known that many characteristics in the "factor zoo" are redundant, and that using a LASSO-type thresholding on the factor premia can effectively improve predictive and pricing performance (e.g., Feng et al., 2020). We take the message of this literature and integrate factor selection into TV-PR, by not only encouraging trend ($\gamma_t = \gamma_{t+1}$) but also sparsity ($\gamma_t = 0$) in selected factors. As a result, TV-PR achieves *time-varying sparsity*

[7] Alternatively, Bandi and Tamoni (2022) and Bandi and Su (2025) rely on spectral methods to decompose returns into different frequency components, and show the intermediate components of business cycle lengths have particular relevance in asset pricing.

[8] Zhu and Zhou (2009) link moving-average indicators to investors' priors. Neely et al. (2014) use trend-type technical indicators to forecast the equity premium. Fama and French (2020) incorporate dynamic characteristics but estimate the model on a period-by-period basis.

in factor selection, allowing the set of active factors to change over time. Thereby, we construct an SDF whose factor composition is not only selected but also varies over time, reflecting changing market conditions and evolving investor preferences.

TV-PR estimates the set of varying risk premia in one step by solving an optimization problem. The problem can be efficiently solved using standard numerical algorithms for the class of LASSO-like optimizations. We employ an Alternating Direction Method of Multipliers (ADMM) algorithm, which yields an estimate in approximately five minutes for the largest model we consider (with over 60 characteristics/factors and around 1 million stock-month observations).

We establish the theoretical properties of the TV-PR estimator. First, the estimator is consistent, ensuring that the estimated factor premia converge to their true regime-dependent values as data accumulate. Second, it is asymptotically normal, permitting valid inference on time-varying risk premia across regimes. Finally, it satisfies the oracle property of break detection, accurately identifying both the relevant predictors and the locations of structural breaks without prior knowledge. In finite sample simulations, TV-PR averages out factor innovations within each regime, resulting in lower estimation errors and higher predictive R^2 values than a range of time-invariant and time-varying benchmark methods.

Empirical findings. We find that breaks and trends are pervasive features of factor premia, and that accounting for these features with the suitable method we propose improves the asset pricing and investment performance, as well as the economic interpretability of the factor models. This overall finding is consistent in different model configurations—using factors one at a time or jointly in a multi-factor model—and applies to both firm-characteristic and past-return factor constructions. We detail these aspects further below.

We perform two groups of empirical analyses. The first one uses 61 firm characteristics to form factors, either individually or jointly in a multi-factor setting. The second narrows down to past returns of different horizons as characteristics in order to inspect the varying premia of momentum, reversal, and seasonality factors. This empirical setting follows Marrow and Nagel (2024), who approach this issue with Gaussian process regressions.

First, we show TV-PR estimates, breaks, and trends in a coherent and economically meaningful way, suggesting that breaks and trends are pervasive and important features in factor premia. We find that the estimated time-varying factor premia tend to have distinct regime shifts around major market events such as the Global Financial Crisis and the COVID-19 pandemic. The patterns of the estimated time-varying premia align with the intuitive trend lines in Figure 1 and in visualizations of the other factors. The identified break points are stable across different rolling estimation windows (provided that the window has adequate margins before and after the break point).[9] For past return-based factors, TV-PR identifies rich decay and recovery patterns in momentum, reversal, and seasonality factors that are consistent with documented findings such as the post-publication decay. It further supports the view that return-based anomalies are not permanent features of the cross-section of expected stock returns (Marrow and Nagel, 2024).

While the estimates conform to the intuitive notion of trends and breaks in factor premia, the next set of results confirms that the trends and breaks identified by TV-PR reflect genuine shifts in the underlying factor premia process, rather than ex post fitting factor realization shocks that happen to align during a continuous period. The distinction is that true factor premia changes are persistent and contain predictive power. The following investment and asset pricing exercises investigate the persistence and predictability properties, which establish the presence of breaks and trends in factor premia as well as their general usefulness.

In the investment exercises, we devise an adaptive factor-timing strategy that uses the time-varying factor premia estimates as trading signals. The rationale is that the most recent factor premia trend should persist and continue into future periods, thereby serving as a timing signal for next month's factor returns. Granted, TV-PR

[9] The TV-PR estimates do not require long margins from the end of the estimation window to detect a new break point, suggesting the potential for "real-time" detection of the breaks (Smith and Timmermann, 2021). This observation also lays the foundation for using trend estimates as factor-timing signals, which is formally detailed and further explained below.

might require several months of observation since the last break to form a reliable estimate of the newest trend, and ”real-time” break detection is always challenging. Nonetheless, so long as the underlying factor premium trend persists after detection, the timing strategy can still be effective. Notice that the TV regularization parameter allows us to explore a spectrum of trend scales ranging from historical averages to the one-month factor momentum signal. Both extremes are degenerate cases of the TV-PR and are well-understood. Somewhere in the middle of the spectrum is intended to provide just the right level of trend-following signals.

We find a hump-shaped relationship between investment performance and TV regularization, indicating a “sweet spot” of trend scales for which the factor-timing strategy is most effective. The peak of the hump shape is stable and dominates the performance of static moving-average-style factor momentum strategies across various window lengths. This result shows the presence of trends in factor premia that persist over time, and that the break-and-trend model of factor premia is particularly suitable when applied to data. It also shows that properly accounting for this structure with TV-PR has significant economic value compared to ad-hoc constructions such as momentum.

The specific strategy design involves monthly rolling-window re-estimates of the TV-PR, where the last (end-of-window) factor premia estimates are used as the trading signal for one month into the future horizon. With that, we perform standard portfolio formation, including single-factor timing strategies and multi-factor long-short strategies, and perform standard backtests and out-of-sample (OOS) evaluations of investment performance, such as Sharpe ratios. The investment performance improvement is consistent across different configurations.

Finally, we evaluate the asset pricing properties of the estimated model, which incorporates breaks and trends in factor premia. The two main asset pricing findings are, first, TV-PR can forecast individual stock returns at the monthly frequency. Statistically, TV-PR can be interpreted as a time-varying coefficient panel predictive regression model, and this finding suggests that the predictive regression framework is working as intended. From an asset pricing perspective, this suggests that factor premia indeed vary with breaks and trends, and that multiplying the premia by the observed factor loadings (or characteristics) can explain the expected stock return in the cross section.

Second, TV-PR enables time-varying factor selection, resulting in an adaptive SDF construction whose factor composition evolves over time. That means not only the factor premia change, but also the composition of the priced factors changes across regimes, reflecting shifts in the sources of systematic risk. We show that the within-regime SDF composition exhibits higher Sharpe ratios compared to a range of classical factor models that remain constant over time (Barillas et al., 2020). It suggests the relevant drivers of cross-sectional returns tend to emerge and fade with economic cycles. Factor models of fixed factor composition are likely mis-specified, and allowing for time-varying factor selection can lead to a more accurate representation of the SDF.

Related literature. Our work contributes to several strands of the literature. First, Smith and Timmermann (2021, 2022) also study break detection methods in factor models. A key methodological distinction we rely on is the machine learning technique of TV regularization, whereas they take a Bayesian approach. The TV specification enables flexible modeling that is closely related to existing methods, such as panel predictive regressions, the Fama-MacBeth procedure, and dynamic factor selection. Moreover, our specification allows for heterogeneous breaks across factors, thereby accommodating distinct structural dynamics that may be masked under a homogeneous specification.[10] Another line of research models time-varying coefficients as functions of observable state variables (Barroso et al., 2021) or scaled time (Chen and Hong, 2012).[11] Our approach is more data-driven and less structured, allowing for

[10]Existing time-varying models often impose the restrictive assumption of simultaneous structural shifts across all predictors. This assumption might overlook the possibility of distinct structural breaks in individual predictors, as noted by Gagliardini et al. (2005), Li et al. (2016), and Smith et al. (2019), thereby failing to capture the heterogeneous dynamics observed in practice.

[11]Moreover, Henkel et al. (2011) employ latent state variables common to all predictors, Pettenuzzo and Timmermann (2011) model time-varying coefficients as rare but large structural breaks in the parameters of return prediction models, Dangl and Halling (2012) follow a random walk within a Bayesian dynamic linear framework, Bonhomme and Manresa (2015) develop a grouped fixed-effects approach for panel data to model the different coefficients across groups, and Yousuf and Ng (2021) introduce

breaks and trends without pre-specifying their functional forms or observable drivers.

This paper also speaks to the literature that addresses the "factor zoo" with selection (e.g., Harvey et al., 2016; Feng et al., 2020). However, most studies assume time-invariant predictive structures with global sparsity, not focusing on the variation in predictor selection over time.[12] TV-PR incorporates time-varying sparsity, permitting predictor importance to evolve over time. We indeed find that time-varying selection improves the construction of SDFs.

This paper contributes to the literature on machine learning predictions (e.g., Gu et al., 2020; Kelly et al., 2019; Bryzgalova et al., 2020), but we present it within a time-varying predictive coefficient framework. Empirically, the predictive relationships are unstable, as Welch and Goyal (2008) highlights that the out-of-sample performance of most equity premium predictors is poor and unstable.[13] We introduce a novel dual-penalization strategy to address prediction instability.

Lastly, this paper contributes to the factor-investment literature by proposing a new factor-timing strategy derived from our estimates of time-varying risk premia. The break-and-trend strategy is closely related to slope factors[14] Fama and French (2020) show that these cross-sectional factors explain average stock returns better than traditional time-series factors. See also Chib et al. (2023) for evidence that slope factors outperform conventional sorted factors, and Kozak and Nagel (2023) for covariance conditions under which they span the SDF. and factor momentum.[15]

This paper is organized as follows. Section 2 introduces the TV-PR estimator, focusing on panel data without prior knowledge of structural changes. Section 3 ex-

time-varying coefficients as a function of time.

[12]Chinco et al. (2019) show that LASSO uncovers sparse, short-lived predictors that improve stock-return forecasts, and Freyberger et al. (2020) use a nonparametric adaptive group LASSO and find that few firm characteristics predict stock returns once nonlinearities are considered.

[13]A key reason for the mixed success of these factors is that their predictive power appears to be time-varying, often concentrated in specific economic regimes or "pockets of predictability" (e.g., Rapach et al., 2010; Farmer et al., 2023; Li et al., 2023).

[14]Since Fama (1976), return–characteristic regression coefficients have served as traded "slope" factors.

[15]Kelly et al. (2021) show that momentum and reversal arise from time-varying risk exposures captured in conditional factor models; Ehsani and Linnainmaa (2022) demonstrate that momentum in individual stock returns relates to momentum in factor returns; and Arnott et al. (2023) find that factor momentum drives industry and stock momentum.

amines its asymptotic properties and finite-sample performance. Section 4 explores time-varying risk premia in a univariate framework. Section 5 extends the model to time-varying predictor selection and develops a return-based predictive approach. Section 6 concludes. Technical proofs and implementation details are in the appendix.

2 Model and TV-PR Estimation

We specify a standard factor model for stock returns, with the only emphasis that the factor premia vary with breaks and trends. Let $r_{i,t}$ denote the excess return of stock i in month t, and $z_{i,t}$ be a $p \times 1$ vector of observed stock characteristics. The factor model is

$$r_{i,t+1} = z_{i,t}' f_{t+1} + u_{i,t+1} \quad \text{with} \quad f_{t+1} = \gamma_t + \nu_{t+1}. \tag{1}$$

Importantly, the factor premia process $\gamma_t \equiv E_t[f_{t+1}]$ is time-varying with trends and breaks; in particular, it follows a piecewise-constant structure:

$$\gamma_{j,1}, \gamma_{j,2}, \gamma_{j,3}, \cdots = \underbrace{\theta_{j,1}, \theta_{j,1}, \ldots, \theta_{j,1}}_{\text{the first trend regime of factor } j}, \underbrace{\theta_{j,2}, \theta_{j,2}, \ldots, \theta_{j,2}}_{\text{the second regime}}, \underbrace{\theta_{j,3}, \theta_{j,3}, \ldots, \theta_{j,3}}_{\text{the third regime}}, \ldots, \tag{2}$$

where the repeated $\theta_{j,n}$ represents the n-th constant-trend segment of factor j's premium. These segments are separated by breaks at unknown locations, which do not need to be aligned across factors. Some of the regimes may have a zero premium ($\theta = 0$), indicating the time-varying sparsity in priced factors. Figure 2a shows this break-and-trend structure in a 10-period three-factor illustration. Section 3 details the assumptions of this data-generating process for deriving the estimator's statistical properties. We also note that the factor model (Eq. 1) takes a simplified view of the factor loadings, which are directly specified as characteristics ($z_{i,t}$). The purpose is to focus our attention on factor premia. From this perspective, model (Eq. 1) is a cross-sectional factor model with time-varying premia.[16]

[16] Cross-sectional factors, also known as slope factors, are factors constructed as the coefficients of period-by-period return-on-characteristics regressions (Fama, 1976; Chib et al., 2023; Fama and French, 2020). They can be interpreted as characteristics-managed portfolio returns, but with improved pricing performances Fama and French (2020). More details can be found in Section 4, where we develop timing strategies based on the observed slope factor returns. A direct slope factor estimation is equivalent to

On the one hand, this model allows time-varying factor premia, unlike most standard factor models (Dai et al., 2019). On the other hand, freely varying premia at each t would be too flexible to estimate. The break-and-trend structure, which is motivated by the economic reality, provides the necessary regularity in how factor premia vary over time, making the model estimable. In particular, the model specifies that breaks are sparse and that trends are constant. The TV-PR estimation method is designed to leverage exactly these two properties.

The TV-PR estimation method is defined as, given a sample observation of excess returns $r_{i,t+1}$ and characteristics $z_{i,t}$ on a panel of $i = 1, \ldots, N$ and $t = 1, \ldots T$, the factor premia estimator ($\hat{\gamma}_t$) solve the following optimization:

$$\min_{\gamma_t \in \mathbb{R}^p} \underbrace{\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(r_{i,t+1} - z_{i,t}' \gamma_t\right)^2}_{\text{predictive regression}} + \underbrace{\lambda \sum_{j=1}^{p} \sum_{t=2}^{T} |\gamma_{t,j} - \gamma_{t-1,j}|}_{\text{total variation}} \; + \; \underbrace{\eta \sum_{j=1}^{p} \sum_{t=1}^{T} |\gamma_{t,j}|}_{\text{factor selection}} . \tag{3}$$

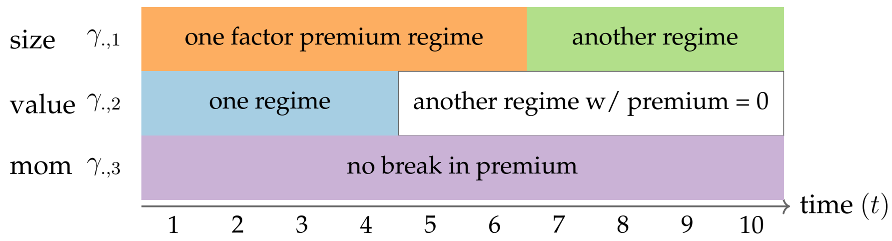
The first term is the standard least squares objective of a panel predictive regression of stock return on characteristics, except that the predictive coefficients are time-varying (γ_t). The second term is TV-PR's key specification that uses total variation regularization to recover breaks and trends in factor premia. It regularizes changes in the γ sequence to induce sparse break points and piecewise-constant trend segments in the $\gamma_{t,j}$ sequence. We use the Lasso-style l_1 regularization to induce exact zero changes, also known as fused Lasso (Tibshirani et al., 2005). The third term is also a penalty term, but is used in a more traditional way. It induces sparsity in the values of γ to facilitate factor selection (rather than the changes of γ).

To better understand TV-PR (Eq. 3), we examine its solution in two extreme cases of the TV regularization (while ignoring factor selection, i.e., $\eta = 0$). A very high TV penalty ($\lambda \to +\infty$) forces γ_t to remain constant over time, and the estimation reduces

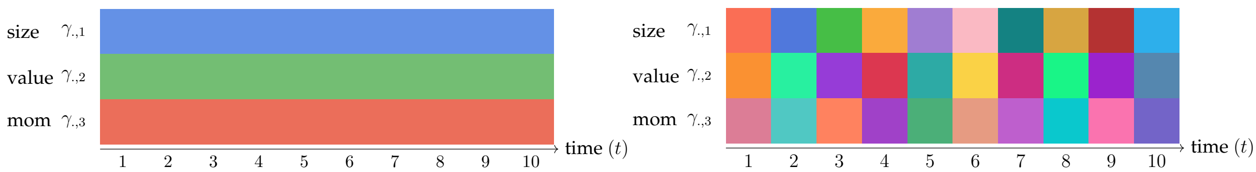
TV-PR extreme case $\lambda = 0$ (no TV regularization) detailed further below. Alternatively, it can also be interpreted as the Barra factor model (Rosenberg, 1974) with break-and-trend factor premia. The simplification of $\beta_{i,t}$ as $z_{i,t}$ avoids the need to estimate time-varying factor loadings, such as in IPCA, where $\beta_{i,t}$ is a linear combination $z_{i,t}$. From this perspective, the first part of (1) is a special case of IPCA, where the number of factors equals the number of characteristics. A joint estimation of time-varying loadings and premia is an interesting direction for future research.

Figure 2: Factor Premia of Different Variability by TV-PR and Its Extreme Cases

Note: Color illustrated made-up samples. Panel 2a illustrates the break-and-trend structure of a possible sample generated by the true model. Factor 1 (say size) has one structural break at $t = 6$, factor 2 has a structural break at a different time, followed by a zero premium regime, and factor 3 has no structural breaks. Panel 2b illustrates the constant factor premia model, as estimated by TV-PR extreme case $\lambda \to +\infty$. Panel 2c illustrates the period-by-period factor premia model, as estimated by TV-PR extreme case $\lambda = 0$.



(a) Breaks and trends in factor premia of the data-generating process



(b) Constant factor premia, case $\lambda \to +\infty$

(c) Period-by-period factor premia, $\lambda = 0$

to the familiar pooled panel predictive regression with fixed coefficients, $r_{i,t+1} = z_{i,t}'\gamma + \epsilon_{i,t+1}$, with

$$\hat{\gamma}_t = \hat{\gamma} = (\sum_{i,t} z_{i,t}z_{i,t}')^{-1} \sum_{i,t} z_{i,t}r_{i,t+1}, \tag{4}$$

where the prediction error $\epsilon_{i,t+1}$ is the composite of factor realization shocks and the stock's idiosyncratic risk: $\epsilon_{i,t+1} = z_{i,t}'\nu_{t+1} + u_{i,t+1}$. Figure 2b illustrates this case. From this perspective, TV-PR with $\lambda < +\infty$ relaxes the constant coefficient assumption and is a dynamic upgrade of factor selection via predictive regressions (e.g., Feng et al., 2020).

On the other hand, without any TV penalty, i.e., $\lambda = 0$, the estimation decouples across t's and reduces to period-by-period cross-sectional regressions of r_{t+1} on z_t, with

$$\hat{\gamma}_t = (z_t'z_t)^{-1}z_t'r_{t+1}. \tag{5}$$

It reduces to the first stage of the Fama-MacBeth two-step procedure, where the regression coefficient is typically interpreted as an estimation of the factor realization ($\hat{f}_{t+1}$). Fama-MacBeth second stage then averages these period-by-period $\hat{f}_{t+1}$ to obtain the factor premia estimate over an sample period, while this TV-PR extreme case directly treats factor realization $\hat{f}_{t+1}$ as the factor premia estimate ($\hat{\gamma}_t$), which is inevitably a noisy premia estimate contaminated by factor realization shocks ν_{t+1}. Figure 2c illustrates this case. From this perspective, TV-PR with $\lambda > 0$ conducts the second-stage averaging of factor realizations ($\hat{f}_{t+1}$) within adaptive segments separated by estimated break points, rather than averaging out ν_{t+1} over arbitrarily specified samples as in rolling-window Fama-MacBeth.

In summary, the innovation of TV-PR lies in drawing a connection between these two well-understood but previously separate procedures and proposing a unified framework that finds a data-adaptive interpolation between them, combining their strengths. The TV regularization allows γ_t to change over time but prevents excessively frequent shifts, thereby capturing both breaks and trends in factor premia. The balancing point of these two forces is controlled by the TV regularization parameter λ, which can be tuned using standard machine learning techniques such as cross-validation. A benefit of TV-PR is that it nests and generalizes these workhorse asset pricing procedures, allowing breaks and trends to be flexibly incorporated into existing techniques, such as factor selection, as we illustrate further below.

3 Statistical Properties

In this section, we formally define the TV-PR estimator for the two cases, with and without prior knowledge of structural breaks, and present a theoretical analysis of the proposed method. Section 3.1 derives its asymptotic properties, and Section 3.2 assesses its finite-sample performance through simulation studies and comparisons with established methods.

We formalize the key features of the time-varying parameter structure in Eq. (1) using mathematical notation. For the j-th candidate predictor, where $j \in \{1, \ldots, p\}$, the coefficient vector $\gamma_{j,\cdot} = (\gamma_{j,1}, \ldots, \gamma_{j,T})'$ can be classified into B_j sets $G_1^j, \ldots, G_{B_j}^j$, such that $G^j_{n_1} \cap G^j_{n_2} = \emptyset$ for any $n_1 \neq n_2$, and $\bigcup_{n=1}^{B_j} G^j_n = \{1, \ldots, T\}$. When $B_j = 1$, the j-th predictor exhibits a homogeneous impact in the econometric model in (1), with coefficient values sharing a constant common value θ^j_1 over time, such that $\gamma_{j,t} = \gamma_{j,\tau} = \theta^j_1$ for all $1 \leq t, \tau \leq T$ and $G^j_1 = \{1, \ldots, T\}$. When $B_j > 1$, there are $B_j - 1$ structural breaks in the j-th predictor's parameter values, with $\gamma_{j,t} = \theta^j_n$, which represents the common coefficient value that the j-th predictor takes when the time period t belongs to the n-th time regime, i.e., $t \in G^j_n$ for $n \in \{1, \ldots, B_j\}$. In sum, we define $B = \sum_{j=1}^p B_j$ as the total number of common values across all regimes, with $B_0 = 0$. Furthermore, we denote $\theta = (\theta_1, \ldots, \theta_B) \in \mathbb{R}^B$ and $\left\{|G^j_n|, j \in \{1, \ldots, p\}, n \in \{1, \ldots, B_j\}\right\} = \{|G_1|, \ldots, |G_B|\}$, where $|G_{n'}| = |G^j_n|$ and $\theta_{n'} = \theta^j_n$ denote the duration and common value parameter for the j-th predictor in its n-th regime G^j_n for $n' = \sum_{m=0}^{j-1} B_m + n$.

To accommodate predictors that may have zero parameter values in certain periods while exhibiting nonzero impacts in others, we introduce a sparsity feature in θ to allow for time-varying sets of useful predictors. We use $\Gamma = (\gamma_1, \ldots, \gamma_T)$ to denote the entire parameter matrix to be estimated. We can partition the corresponding common value parameter vector θ into two subvectors: $\theta = (\theta'_{\mathbb{A}}, \theta'_{\mathbb{Z}})'$, where $\theta_{\mathbb{A}} \in \mathbb{R}^q$ contains the nonzero elements and $\theta_{\mathbb{Z}} \in \mathbb{R}^{B-q}$ contains the zero elements of θ.

Without any prior knowledge on structural changes and the usefulness of predictors, we consider obtaining $\hat{\Gamma}$ from the following generalized double-penalized panel predictive regression procedure:

$$\min_{\gamma_t \in \mathbb{R}^p} \frac{1}{NT} \sum_{i,t} \left(r_{i,t+1} - z'_{i,t}\gamma_t\right)^2 + \sum_{j=1}^{p} \sum_{t=2}^{T} P_\lambda(|\gamma_{t,j} - \gamma_{t-1,j}|) + \sum_{j=1}^{p} \sum_{t=1}^{T} P_\eta(|\gamma_{t,j}|). \quad (6)$$

where we generalize the penalty functions to allow for a general folded concave penalty function for P_η and P_λ. For example, it includes the smooth clipped absolute derivatives (SCAD) proposed by Fan and Li (2001), which is defined as follows:

$$P_\lambda(x) = \lambda|x|I(0 \leq |x| < \lambda) + \frac{a\lambda|x| - (x^2 + \lambda^2)/2}{a-1} I(\lambda \leq |x| \leq a\lambda) + \frac{(a+1)\lambda^2}{2} I(|x| > a\lambda),$$

where $a > 2$. The same specifications apply to $P_\eta(x)$.[17]

[17] When employing the minimax concave penalty (MCP) in Zhang (2010), $P'_\lambda(x) = sign(x)[\lambda - |x|/a]$ if $|x| \leq a\lambda$, and 0 otherwise, where $a > 1$.

We identify a significant relationship between the Lasso penalty regularized TV-PR estimator, as presented in (3), and the estimator in (6) that utilizes a general folded concave penalized regularization.[18] First, folded concave penalty functions are effective in handling correlated covariates, which can complicate the irrepresentable condition typically required by standard Lasso penalty functions. However, this advantage comes with a trade-off: the use of folded concave penalties introduces nonconvexity, leading to the possibility of multiple local solutions.

To achieve a global solution for (6) while retaining the broad applicability of folded concave penalty functions, we propose employing the local linear approximation (LLA) method as described by Zou and Li (2008). The LLA algorithm transforms a nonconvex regularization problem into a series of weighted l_1-penalized problems. Consequently, we can utilize (3), the Lasso-regularized TV-PR estimator, as the initial estimator $\hat{\Gamma}^{\text{initial}}$and obtain LLA through iterations as follows:

$$\min_{\gamma_t \in \mathbb{R}^p} \frac{1}{NT} \sum_{i,t} \left(r_{i,t+1} - z'_{i,t}\gamma_t\right)^2 + \sum_{j=1}^{p}\sum_{t=2}^{T} \hat{w}_{t,j,\lambda}^{(m-1)} |\gamma_{t,j} - \gamma_{t-1,j}| + \sum_{j=1}^{p}\sum_{t=1}^{T} \hat{w}_{t,j,\eta}^{(m-1)} |\gamma_{t,j}|, \quad (7)$$

where $\hat{w}_{t,j,\lambda}^{(m-1)} = P'_\lambda(|\hat{\gamma}_{t,j}^{(m-1)} - \hat{\gamma}_{t-1,j}^{(m-1)}|)$ and $\hat{w}_{t,j,\eta}^{(m-1)} = P'_\eta(|\hat{\gamma}_{t,j}^{(m-1)}|)$, and m is the number of iterations of LLA.

Hence, such a numerical procedure integrates the desirable properties from (3) and (6) and ultimately leads to a unique global minimizer. We provide a detailed description of the numerical implementation algorithm in the online Appendix.

Discussion: Prior information on the timing of structural breaks. Now, we investigate the role that prior information of structural changes can play in the estimation of time-varying risk premia. When econometricians obtain some prior information on the timing of regimes, they can construct a selection matrix M that reflects the information on structural breaks and bridges the parameter matrix and common value vector where $vec(\Gamma) = M\theta$.[19]

[18] The theoretical justifications for these penalty functions will be discussed in Assumption 1.

[19] $vec(\Gamma) = (\gamma'_1, \dots, \gamma'_T)'$ and $M = (M'_1, \dots, M'_T)'$ is the $pT \times B$ selection matrix with M_t being a $p \times B$ 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

We emphasize an intriguing finding: when prior information about structural breaks is available, our proposed double-penalized TV-PR estimator can be reduced to an equivalent weighted Lasso regularized estimator, as detailed below. With prior knowledge of the timing of structural changes, the estimates of time-varying parameter values in (6) are simplified to

$$\min_{\theta\in\mathbb{R}^B} \frac{1}{NT}\|Y - \mathbf{Z}\theta\|^2 + \sum_{n=1}^{B} |G_n| P_\eta(|\theta_n|), \tag{8}$$

where $P_\eta(\cdot)$ is a folded concave penalty function with a regularization parameter η, $\|\cdot\|$ is the l_2 norm of a vector, and we employ the duration of each time region $|G_n|$ as the heterogeneous penalty weights in the model. For notational convenience, we define $r_t = (r_{1,t}, \ldots, r_{N,t})'$ as the dependent variable, $Z_t = (z_{1,t}, \ldots, z_{N,t})'$ as an $N \times p$ matrix, and $\epsilon_t = (\epsilon_{1,t}, \ldots, \epsilon_{N,t})'$ for $t = 1, \ldots, T$. We denote $Y = (r_2', \ldots, r_{T+1}') \in \mathbb{R}^{NT}$ as the vector of dependent observations, $Z = (Z_1', \ldots, Z_T')$ as a $p \times NT$ matrix, and $\tilde{Z} = \text{diag}(Z_1, \ldots, Z_T)$. Denote $\mathbf{Z} = \tilde{Z}M = (\mathbf{Z}_1^1, \ldots, \mathbf{Z}_{B_1}^1, \ldots, \mathbf{Z}_1^p, \ldots, \mathbf{Z}_{B_p}^p)$ as a $NT \times B$ data matrix, where $\mathbf{Z}_s^j$ is a $NT \times 1$ vector.[20]

3.1 Large Sample Properties

We will first discuss the asymptotic guarantees of our TV-PR estimator, considering both scenarios: with and without prior knowledge of structural breaks. We introduce some additional notations. Let $\lambda_{\min}(D)$, $\lambda_i(D)$ and $\lambda_{\max}(D)$ denote the minimum, the i-th, and the maximum eigenvalue of a $K \times K$ matrix D. For a $K_1 \times K_2$ matrix D, denote $\|D\|_{2,\infty} = \max_{1\le k\le K_1} \|D_{k,}\|$ with $\|D_{k,}\| = \sqrt{\sum_{1\le j\le K_2} D_{kj}^2}$. Let $\mathbb{S}_G = \{\Gamma \in \mathbb{R}^{T\times p} : \gamma_{j,t} = \gamma_{j,\tau}, \text{for } t, \tau \in G_n^j, 1 \le n \le B_j, j = 1, \ldots, p\}$. Let $D_{\mathbb{A}} = diag(|G_1|^{1/2}, \ldots, |G_q|^{1/2})$. Denote $\epsilon_{t+1} = z_t'\nu_{t+1} + u_{t+1}$ and define $\boldsymbol{\epsilon} = (\epsilon_2', \ldots, \epsilon_{T+1}')$, which is a $NT \times 1$vector. Given the separation of $\mathbf{Z} = (\mathbf{Z}_{\mathbb{A}}, \mathbf{Z}_{\mathbb{Z}})$ according to $\theta = (\theta_{\mathbb{A}}', \theta_{\mathbb{Z}}')'$, we further define $\Phi_{\mathbb{A}} = \frac{1}{N}E[D_{\mathbb{A}}^{-1}\mathbf{Z}_{\mathbb{A}}'\mathbf{Z}_{\mathbb{A}}D_{\mathbb{A}}^{-1}]$ and $\Sigma_{\mathbb{A}} = \frac{1}{N}var[D_{\mathbb{A}}^{-1}\mathbf{Z}_{\mathbb{A}}'\boldsymbol{\epsilon}]$. We denote $\mathbf{Z}_{\mathbb{A},i}$ as the i-th row of $\mathbf{Z}_{\mathbb{A}}$ and $\mathbf{Z}_{\mathbb{A},\cdot,j}$ as the j-th column of $\mathbf{Z}_{\mathbb{A}}$. Similar notation applies to $\mathbf{Z}_{\mathbb{Z},i}$ and $\mathbf{Z}_{\mathbb{Z},\cdot,k}$ with

for all $t \in \{1, \ldots, T\}$.

[20] If $t \in G_s^j$, the elements from the $(\sum_{i=0}^{t-1} N_i + 1)$-th to $\sum_{i=0}^{t} N_i$-th locations equal $z_{\cdot,t,j}$; otherwise, the elements are zeros.

$i = 1, \ldots, NT$, $j = 1, \ldots, q$, and $k = 1, \ldots, B - q$.

Assumption 1. *The penalty function $P_\lambda(x)$ is symmetric, nondecreasing, and concave in $x \in [0, \infty)$, and $P_\lambda(0) = 0$. For some $a > 0$, $P_\lambda(x)$ is a constant for $x > a\lambda$. $P'_\lambda(x)$ is increasing in $\lambda \in [0, \infty)$ and $P'_\lambda(x)$ is nonincreasing in $x \in [0, \infty)$ with $\lambda^{-1} P'_\lambda(0+) = a_1$ and $P'_\lambda(x) \geq a_1 \lambda$ for $x \in (0, a_2\lambda]$. The same requirements hold for $P_\eta(x)$.* [21]

Assumption 2. *Assume $\{(z'_{i,1}, \ldots, z'_{i,T}, u'_i)\}_{i=1}^N$ is an independent sequence. Suppose that for all $i = 1, \ldots, N$, $j = 1, \ldots, p$, and $t = 1, \ldots, T$, $E[u_{i,t} | r_{i,t-1}, \ldots, r_{i,1}, z_{i,t}, \ldots, z_{i,1}] = 0$ and $E[\nu_t | r_{i,t-1}, \ldots, r_{i,1}, z_{i,t}, \ldots, z_{i,1}] = 0$.*

Assumption 3. *Suppose there exist constants $C_1, C_2 > 0$ such that, for every $\kappa \geq 0$, $t = 1, \ldots, T$, and $i = 1, \cdots, N$, $\sup_{\|a\| \leq 1} P(|a'\nu_t| \geq \kappa) \leq C_1 e^{-C_2 \kappa^2}$ and $\sup_{\|b\| \leq 1} P(|b' u_t| \geq \kappa) \leq C_1 e^{-C_2\kappa^2}$. Moreover, $z_{i,t}$ are uniformly subgaussian that $\sup_{\|e\| \leq 1} P(|e' z_{i,t}| > \kappa) \leq C_1 e^{-C_2 \kappa^2}$.*

Assumption 4. *Suppose (i) the eigenvalues of the $q \times q$ matrix $E[\frac{1}{N} \mathbf{D}_{\mathbb{A}}^{-1} \mathbf{Z}'_{\mathbb{A}} \mathbf{Z}_{\mathbb{A}} \mathbf{D}_{\mathbb{A}}^{-1}]$ are bounded below by $C_3 > 0$ and bounded above by $C_4 > 0$. (ii) the eigenvalues of $\Sigma_{\mathbb{A}} = \frac{1}{N} var[D_{\mathbb{A}}^{-1} \mathbf{Z}'_{\mathbb{A}} \boldsymbol{\epsilon}]$ are bounded below by $C_5 > 0$ and bounded above by $C_6 > 0$. (iii) There exists positive constants C_7 and C_8 such that for all $i = 1, \ldots, q$ and $j = 1, \ldots, NT$, $\sup_i E\|\mathbf{Z}_{\mathbb{A},i}\|^2 \leq C_7 N G_i$, $\sup_j E\|\mathbf{Z}_{\mathbb{Z},\cdot,j}\|^2 = C_8 q$, and $E\|\mathbf{Z}'_{\mathbb{Z}} \mathbf{Z}_{\mathbb{A}}\|_{2,\infty} = O(NT)$.*

Assumption 5. *Suppose (i) $P'_\eta(d_T) = O\big(\sqrt{G_{\min}}/(T|G_{\max}|)\big)$ with $d_T \geq q/G_{\min}$. (ii) Let $\mathbf{N}_0 = \{\Theta \in \mathbb{R}^q : \|\Theta - \theta_{o\mathbb{A}}\| \leq d_T\}$, for $\theta \in \mathbf{N}_0$, $\max_n P''_\eta(\Theta_n) = o\big(|G_{\min}|/|G_{\max} T|\big)$. (iii) $P'_\eta(0+) \geq \max\{\sqrt{q/|G_{\min}|}/|G_{\min}|, \sqrt{\ln(NG_{\min})/NG_{\min}^{1-\alpha}}/|G_{\min}|\}$. (iv) $\ln(B) = O((NG_{\min})^\alpha)$ for some $\alpha \in (0, 1/2)$. (vi) $\lambda \geq c_8\big[\big(\sqrt{q\ln(NT)/NT} + \eta + \eta|G_{\min}|\big)\big]/G_{\min}]$ for some positive constant c_8.*

Assumption 1 states formal requirements for penalty functions, enabling us to circumvent irrepresentable conditions and accommodate correlated candidate covariates. Assumption 2 imposes independence across individuals $i = 1, \ldots, N$, a common condition in panel data analysis.[22] We relax the assumption on temporal dependence,

[21]For SCAD, $a_1 = a_2 = 1$, $P'_\lambda(x) = \text{sign}(x)\lambda$ if $|x| \leq \lambda$, and $\text{sign}(x)\frac{(a\lambda - |x|)_+}{a-1}$ otherwise. For MCP in Zhang (2010), $P'_\lambda(x) = \text{sign}(x)\,[\lambda - |x|/a]$ if $|x| \leq a\lambda$, and 0 otherwise, where $a_1 = 1 - 1/a$ and $a_2 = 1$.

[22]However, this assumption can be relaxed to allow for cross-sectional dependence by incorporating latent structures, as demonstrated in Su and Wang (2017).

such as strong mixing conditions, widely used in the econometric literature, but at a cost of involving $\ln(p \vee N)^3$, rather than $\ln(p \vee N)$ in the penalty terms η and λ. Importantly, we do not assume identical error distributions across individuals or time. Additionally, conditional heteroskedasticity in the error terms is permitted. Assumption 3 pertains to the restricted eigenvalue condition discussed in Bickel et al. (2009), a standard assumption in the variable selection literature. This condition is naturally satisfied when Φ is positive definite. Assumption 4 assumes that both the error terms and predictors are uniformly subgaussian. The general applicability of this assumption in high-dimensional panel data regression models has been illustrated by Kock and Tang (2019). Assumption 5(i) also relates to the restricted eigenvalue condition as outlined in Bickel et al. (2009). Furthermore, it is employed by Ke et al. (2015) to capture homogeneity among regressors.

Theorem 1 (Consistency)**.** *Suppose Assumptions 1-5 hold. A strict local minimizer* $\tilde{\theta} = (\tilde{\theta}'_{\mathbb{A}}, \tilde{\theta}'_{\mathbb{Z}})'$ *of (8) exists and* $\|\tilde{\theta} - \theta\| = O_p\big(\sqrt{q/NG_{\min}}\big)$.

Theorem 1 examines the estimation consistency of our proposed method. It is important to note that the convergence rate of our method is influenced by the cross-sectional dimension N, the minimum duration across all regimes $G_{\min}$, and the total number of effective regimes q. There is a clear trade-off among these dimensions.

Now, we discuss how to perform inference for $\tilde{\theta}$. For $\Sigma_{\mathbb{A}}$, we have $\Sigma_{\mathbb{A},n'n'} = \frac{1}{N|G_n^j|}\sum_{i=1}^N \sum_{t\in G_n^j} E[z_{i,t,j}\epsilon_{i,t+1}]^2$ and $\Sigma_{\mathbb{A},n'v'} = \frac{1}{N\sqrt{|G_n^j||G_v^w|}}\sum_{i=1}^N \sum_{t\in G_n^j \cap G_v^w} E[z_{i,t,j}z_{i,t,w}\epsilon_{i,t+1}^2]$, where $n' = \sum_{l=1}^{j-1} B_l + n$ and $v' = \sum_{l=1}^{w-1} B_l + v$ with $1 \le n \le B_j$ and $1 \le v \le B_w$ for $1 \le j, w \le p$. We construct a feasible sample counterpart of $\Sigma_{\mathbb{A}}$ as $\tilde{\Sigma}_{\mathbb{A}}$ such that $\tilde{\Sigma}_{\mathbb{A},n'n'} = \frac{1}{N|G_n^j|}\sum_{i=1}^N \sum_{t\in G_n^j} \tilde{\epsilon}_{i,t+1}^2 z_{i,t,j}^2$ and $\tilde{\Sigma}_{\mathbb{A},n'v'} = \frac{1}{N\sqrt{|G_n^j|\sqrt{|G_v^w|}}}\sum_{i=1}^N \sum_{t\in G_n^j \cap G_v^w} \tilde{\epsilon}_{i,t+1}^2 z_{i,t,j}z_{i,t,w}$, where $\tilde{\epsilon}_{i,t+1} = r_{i,t+1} - z'_{i,t}\tilde{\gamma}_t$, where $\tilde{\Gamma} = (\tilde{\gamma}_1, \ldots, \tilde{\gamma}_T)$ such that $vec(\tilde{\Gamma}) = M\tilde{\theta}$ with $\tilde{\theta}$ that solves (8). Consider a feasible sample counterpart of $\Phi_{\mathbb{A}}$ as $\tilde{\Phi}_{\mathbb{A}}$, where $\tilde{\Phi}_{\mathbb{A},n'v'} = 0$ if $G^j \cap G_v^w = \emptyset$, $\tilde{\Phi}_{\mathbb{A},n'n'} = \frac{1}{N|G_n^j|}\sum_{i=1}^N \sum_{t\in G_n^j} z_{i,t,j}^2$, and $\tilde{\Phi}_{\mathbb{A},n'v'} = \frac{1}{N\sqrt{|G_n^j||G_v^w|}}\sum_{i=1}^N \sum_{t\in G_n^j \cap G_v^w} z_{i,t,j}z_{i,t,w}$.

Theorem 2 (Asymptotic Normality)**.** *Suppose Assumptions 1-5 hold and assume* $P'_\eta(d_T) = o\big(|G_{\min}|/(\sqrt{qTN}|G_{\max}|)\big)$ *with* $d_T = \min_{1\le n\le q}|\theta_{\mathbb{A},n}|/2$ *as half of the minimum signal. For* *some unit vector* $\zeta \in \mathbb{R}^q$ *with* $\|\zeta\|^2 = 1$*, we have*

$$\frac{\zeta'\sqrt{N}D(\tilde{\theta}_{\mathbb{A}} - \theta_{\mathbb{A}})}{\sqrt{\zeta'\tilde{\Phi}_{\mathbb{A}}^{-1}\tilde{\Sigma}_{\mathbb{A}}\tilde{\Phi}_{\mathbb{A}}^{-1}\zeta}} \xrightarrow{d} N(0,1),$$

with

$$|\zeta'\tilde{\Phi}_{\mathbb{A}}^{-1}\tilde{\Sigma}_{\mathbb{A}}\tilde{\Phi}_{\mathbb{A}}^{-1}\zeta - \zeta'\Phi_{\mathbb{A}}^{-1}\Sigma_{\mathbb{A}}\Phi_{\mathbb{A}}^{-1}\zeta| = o_p(1).$$

Moreover

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

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

Theorem 2 provides inference on the performance of a single regime when $\zeta_i = 1$ for the i-th regime and zero otherwise. When $\zeta_i = \frac{1}{\sqrt{q}}$, Theorem 2 assesses whether slope factors have a zero predictive ability on average. Theorem 2 provides theoretical assurances for selection consistency.

Theorem 3 (Oracle Property)**.** *Suppose Assumptions 1-5 hold. A positive constant* c *exists whereby* $\min_{1\leq n\leq q}|\theta_{\mathbb{A},n}| > c\eta$*. Let* $\omega_T = \min_{1\leq j\leq p, N_j>1}\min_{n_1\neq n_2}|\theta^j_{n_1} - \theta^j_{n_2}|$ *and* $\omega_T > c\lambda$*, with* $\lambda \gg 1/\sqrt{T}$*. Suppose* $\lambda > [c_1\sqrt{q\ln(NT)/NT} + \eta|G_{\min}| + \eta]/|G_{\min}|$*. A local minimizer* $\hat{\Gamma}$ *of the penalized criterion (6) exists such that*

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

where $\tilde{\Gamma}^*$ *is the oracle estimator corresponding to* $\tilde{\theta}^* = (\tilde{\theta}^{*\prime}_{\mathbb{A}}, \mathbf{0}_{B-q})'$ *that* $vec(\tilde{\Gamma}^*) = M\tilde{\theta}^*$*, which minimizes the following moment conditions:* $\tilde{\theta}^*_A = \arg\min_{\theta_A\in\mathbb{R}^q}\frac{1}{NT}\sum_{t=1}^{T}\|Y - \mathbf{Z}_{\mathbb{A}}\theta_{\mathbb{A}}\|^2$.

Theorem 3 offers statistical guarantees for our estimator in (6) when we do not rely on prior information regarding structural changes or asset return regimes, thereby ensuring its robust application in a general cross-sectional regression framework.

3.2 Finite Sample Properties

In this section, we evaluate the finite-sample performance of the TV-PR estimator and benchmark it against established methods across varying sample sizes and cross-sectional dimensions. The analysis utilizes a calibrated data-generating process (DGP) based on empirical data.

We simulate a three-regime, three-factor specification of model (1), using data-generating parameters calibrated to empirical factor and stock returns. Specifically, we set

$$r_{i,t+1} = z_{i,t}^{[1]} f_{t+1}^{[1]} + z_{i,t}^{[2]} f_{t+1}^{[2]} + z_{i,t}^{[3]} f_{t+1}^{[3]} + u_{i,t+1}, \quad t = 1, \dots, T, \tag{9}$$

where each factor is generated with time-varying, piecewise-constant factor premia, featuring structural breaks at $T/3$ and $2T/3$. In particular, we define $f_{t+1}^{[k]} = \gamma_t^{[k]} + \nu_{t+1}^{[k]},\ k = 1, 2, 3$, where $\nu_{t+1}^{[k]}$ varies across the three regimes.[23] We simulate both factor innovation as ν_{t+1} and idiosyncratic errors as i.i.d. Normal variables with calibrated standard deviations.[24] The predictors are generated from a VAR(1) process: $z_{i,t} = Az_{i,t-1} + e_{i,t},\ e_{i,t} \sim N(0, \Sigma)$.[25]

We experiment with four sample size configurations to assess the finite-sample performance of TV-PR with $T = 120$ or 240 and $N = 500$, or 1000. For each configuration, we conduct 50 simulations to evaluate the estimation accuracy. To assess the finite-sample performance of our method, we compare TV-PR with several alternatives, including time-invariant model selection methods (LASSO and Ridge), a standard time-invariant estimator (OLS), and time-varying approaches based on a 120-month rolling average and a recursive average. We evaluate estimation accuracy us-

[23]Specifically, the factor premia are specified as the $\gamma_t^{[k]} = (1 + E_t^{\text{ann}}[f_{t+1}^{[k]}])^{1/12} - 1$, where $E_t^{\text{ann}}[f_{t+1}^{[k]}]$ denotes the annual expected factor return, following: $E_t^{\text{ann}}[f_{t+1}^{[1]}] = 10\%,\ E_t^{\text{ann}}[f_{t+1}^{[2]}] = 8\%,\ E_t^{\text{ann}}[f_{t+1}^{[3]}] = -3\%$, for $t = 1, \dots, T/3$; $E_t^{\text{ann}}[f_{t+1}^{[1]}] = -10\%,\ E_t^{\text{ann}}[f_{t+1}^{[2]}] = 3\%,\ E_t^{\text{ann}}[f_{t+1}^{[3]}] = 6\%$, for $t = T/3 + 1, \dots, 2T/3$; $E_t^{\text{ann}}[f_{t+1}^{[1]}] = 5\%,\ E_t^{\text{ann}}[f_{t+1}^{[2]}] = -5\%,\ E_t^{\text{ann}}[f_{t+1}^{[3]}] = 12\%$, for $t = 2T/3 + 1, \dots, T$.

[24]$\nu_{t+1} \sim$ i.i.d. $N(0, \sigma_\nu^2)$, with $\sigma_\nu = 0.29\%$. The standard deviation of the idiosyncratic errors is calibrated such that the total R^2 of Eq. (1) is approximately 20%. The total R^2 is defined as $1 - \sum_{i,t} \varepsilon_{i,t+1}^2 / \sum_{i,t} r_{i,t+1}^2$.

[25]Matrices A and Σ are estimated with empirical characteristics. For each i, the sequence of $z_{i,t}$ is generated from a separate VAR process, discarding the first ten realizations.

ing two performance metrics: (i) average absolute estimation error (Abs.Err.), defined as $(pT)^{-1}\sum_j^p \sum_t^T |\hat{\theta}_{j,t} - \theta_{j,t}|$; and (ii) the predictive R^2, defined as $1 - \frac{\sum_{i,t}(r_{i,t+1} - \hat{r}_{i,t+1})^2}{\sum_{i,t} r_{i,t+1}^2}$.

Figure 3: Estimation of Risk Premia from 50 Simulations

Note: This figure illustrates the estimation of risk premia across 50 simulations with $N = 1000$ and $T = 240$. The transparent solid lines represent the estimated risk premia, while the dashed lines indicate their true values.

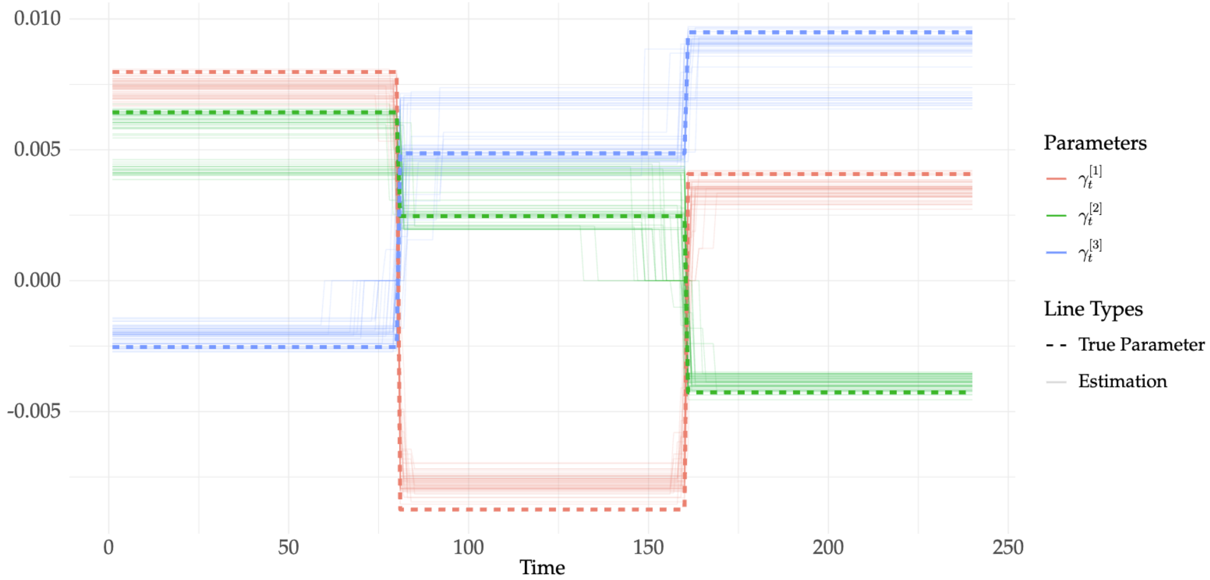


Figure 3 presents the true risk premia and their estimates from 50 simulations ($N = 1000$ and $T = 240$). The dashed lines represent the true risk premia, while the solid lines depict the estimates from each simulation, three three colors corresponding to the three factors, respectively.

Overall, the simulation results validate the effectiveness of our time-varying risk premia estimation framework. In most simulations, the estimated risk premia closely track the true values, suggesting the factor innovations within each regime are effectively averaged out to reveal the underlying risk premia.[26] The timing of breaks aligns well with the true breaks, occasionally appearing earlier or later than the true structural breaks. Nonetheless, these deviations are modest, and the method successfully recovers the true regime structure in most cases. These findings highlight the method's

[26]The Fama–MacBeth regression conducts period-by-period cross-sectional regressions and then averages the estimated coefficients over time to obtain the risk premia. This time-series averaging helps average out factor innovations in the estimated premia, as such innovations can induce substantial estimation errors (Giglio and Xiu, 2021). Our method requires a minimum regime duration to ensure that factor innovations are sufficiently averaged out within each regime.

suitability for empirical applications.

However, some estimation errors and biases are observed and behave in expected ways according to TV-PR's design. First, in some simulations, the green lines across regimes 1 and 2 are fused together. This occurs as the total variation penalty fails to detect the first break point in factor 2's risk premia, given the relatively small magnitude of change. In such cases, the estimates tend to converge to an average of the true risk premia across the adjacent regimes, reflecting the difficulty of detecting weak structural breaks. Third, the risk premium estimates exhibit mild shrinkage toward zero, a direct consequence of the Lasso penalty. This regularization effect is particularly beneficial in high-dimensional settings, as it guards against overfitting and enhances out-of-sample stability, albeit at the cost of introducing slight bias. For the same reason, in a few simulations, the factor premia are erroneously set to zero in between two regimes where the premia switch signs (blue lines around $T = 80$ and green lines around $T = 160$).

Table 1: Average Absolute Estimation Error of Risk Premia

Note: This table compares the estimation accuracy of different methods across two time horizons ($T = 120$ and $T = 240$) and two sample sizes ($N = 500$ and $N = 1000$). Coefficient precision is evaluated using the average absolute estimation error (Abs.Err), expressed as a percentage. Results are averaged over 50 simulations, with standard deviations in parentheses.

		$N = 500, T = 120$				$N = 500, T = 240$		
	TV-PR	Lasso	Ridge	OLS	TV-PR	Lasso	Ridge	OLS
All Regimes	0.07	0.49	0.49	0.49	0.08	0.49	0.49	0.49
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.00)	(0.00)	(0.00)
		$N = 1000, T = 120$				$N = 1000, T = 240$		
	TV-PR	Lasso	Ridge	OLS	TV-PR	Lasso	Ridge	OLS
All Regimes	0.07	0.49	0.49	0.49	0.08	0.49	0.49	0.49
	(0.02)	(0.01)	(0.01)	(0.01)	(0.03)	(0.00)	(0.00)	(0.00)

Table 1 presents estimation accuracy across sample sizes, highlighting the robustness of our method relative to alternatives. TV-PR achieves average absolute estimation errors approximately seven times lower than those of competing methods, demonstrating superior precision in parameter estimation. In contrast, time-invariant estimation methods can only estimate the time-invariant risk premia as the average across all regimes, thereby inducing large estimation errors.

Table 2: Predictive R^2 across Different Models

Note: This table compares model predictability across different methods for two time horizons ($T = 120$ and $T = 240$) and two sample sizes ($N = 500$ and $N = 1000$). Predictability is evaluated using the predictive R^2, reported in percentages. Results are averaged over 50 simulations, with standard deviations in parentheses.

	$N = 500, T = 120$						$N = 500, T = 240$					
	TV-PR	Lasso	Ridge	OLS	Rolling	Recursive	TV-PR	Lasso	Ridge	OLS	Rolling	Recursive
Regime 1	17.82	-0.64	-0.56	-0.61	0.21	1.04	17.64	-0.61	-0.53	-0.58	0.35	1.65
	(1.68)	(1.41)	(1.39)	(1.41)	(1.04)	(1.08)	(1.32)	(0.97)	(0.95)	(0.97)	(0.72)	(0.56)
Regime 2	16.46	1.98	1.96	1.94	-4.83	-2.72	16.28	2.06	2.04	2.02	-4.11	-1.90
	(1.17)	(1.15)	(1.14)	(1.15)	(0.82)	(0.80)	(1.02)	(0.81)	(0.80)	(0.81)	(0.56)	(0.38)
Regime 3	19.48	7.65	7.59	7.66	3.38	0.16	19.33	7.71	7.65	7.72	4.11	0.22
	(1.55)	(1.24)	(1.22)	(1.24)	(1.33)	(0.69)	(1.30)	(0.96)	(0.95)	(0.96)	(0.88)	(0.41)
All	17.96	3.05	3.05	3.05	-0.35	-0.48	17.78	3.09	3.09	3.09	0.17	0.00
	(0.81)	(0.45)	(0.45)	(0.45)	(0.67)	(0.29)	(0.71)	(0.32)	(0.32)	(0.32)	(0.42)	(0.13)
	$N = 1000, T = 120$						$N = 1000, T = 240$					
	TV-PR	Lasso	Ridge	OLS	Rolling	Recursive	TV-PR	Lasso	Ridge	OLS	Rolling	Recursive
Regime 1	17.62	-0.79	-0.71	-0.77	0.08	0.86	17.54	-0.72	-0.64	-0.69	0.30	1.61
	(1.50)	(1.19)	(1.17)	(1.19)	(0.94)	(0.94)	(1.14)	(0.75)	(0.74)	(0.75)	(0.60)	(0.44)
Regime 2	16.48	2.02	2.00	1.98	-4.71	-2.47	16.22	2.08	2.06	2.04	-4.07	-1.94
	(1.21)	(0.98)	(0.97)	(0.98)	(0.79)	(0.63)	(0.96)	(0.82)	(0.81)	(0.82)	(0.55)	(0.34)
Regime 3	19.43	7.69	7.62	7.70	3.38	0.00	19.26	7.68	7.61	7.69	4.11	0.32
	(1.58)	(1.30)	(1.28)	(1.31)	(1.20)	(0.64)	(1.14)	(0.81)	(0.80)	(0.81)	(0.65)	(0.26)
All	17.88	3.02	3.02	3.02	-0.36	-0.52	17.70	3.05	3.05	3.05	0.17	0.01
	(0.79)	(0.39)	(0.39)	(0.39)	(0.56)	(0.22)	(0.64)	(0.28)	(0.28)	(0.28)	(0.30)	(0.11)

Table 2 reports the predictive R^2 across different methods. First, TV-PR consistently achieves the highest predictive R^2 across the three regimes, whereas the time-invariant methods exhibit mixed performance across regimes, highlighting their limited ability to capture dynamic relationships. This demonstrates TV-PR's reliability in producing accurate predictions under varying conditions and structural breaks. Second, the time-varying estimation methods, rolling and recursive averages, perform surprisingly worse than the time-invariant methods, as these two approaches can only capture gradual changes rather than structural breaks. However, whether rolling or breaks perform better depends on the true DGP. Determining which approach is more applicable in real-world settings is an empirical question. Later, we show that the break-based method indeed performs better, suggesting that the true DGP of financial data is characterized more by structural breaks than by gradual rolling changes.

4 TV-PR with A Single Characteristic

We begin by showing, in a univariate setting, that accounting for both structural breaks and trend dynamics in factor premia is essential. The investment strategy based on time-varying risk premia outperforms the ad hoc break-trend factor-momentum strategy, and the evidence shows that the breaks and trends identified by TV-PR reflect genuine shifts in the underlying factor-premia process. Building on this evidence, we extend the analysis in the next section to a multivariate framework that incorporates time-varying factor models and forecast-implied portfolio allocations, after which we revisit return-based anomalies.

4.1 Data

Our analysis uses a standard dataset of US stock characteristics and returns. We use monthly U.S. equity data spanning 2004 to 2023. We apply data filters consistent with those used in constructing the Fama-French factors. Specifically, the sample is restricted to stocks listed on NYSE, AMEX, or NASDAQ for at least one year, with CRSP share codes of 10 or 11. Firms with negative book equity or lagged market equity are excluded to maintain data integrity. The dataset encompasses 61 firm characteristics observed monthly, categorized into 8 clusters: momentum, value, investment, profitability, size, volatility, liquidity, and intangibles. These characteristics are standardized to the interval $[-1, 1]$ on a cross-sectional basis.

4.2 Breaks and Trends Estimation in Factor Premia

The factor premia may experience structural breaks across regimes, as documented by Henkel et al. (2011) and Smith and Timmermann (2021), while also displaying varying trend dynamics across factors. To analyze these patterns, we first employ univariate analysis, which offers a clear framework for detecting structural breaks and identifying trends.[27] Specifically, we focus on cross-sectional factors $\hat{f}_{t+1}$, which estimate the latent factors f_{t+1} from Eq. (1) as

$$\hat{f}_{t+1} = (Z_t' Z_t)^{-1} Z_t' r_{t+1} = \gamma_t + (Z_t' Z_t)^{-1} Z_t' \epsilon_{t+1}, \tag{10}$$

where $\epsilon_{t+1} = Z_t \nu_{t+1} + u_{t+1}$ is the composite of factor innovations and idiosyncratic shocks. In this view, the purpose of TV-PR is to recover the breaks and trends in

[27] We retain a constant and one characteristic in Eq. (1), applying the total variation penalty in Eq. (6). This approach streamlines the model while preserving its essential structure.

the observed slope factors $\hat{f}_{t+1}$. When Z_t includes a column of ones, the intercept estimate $\hat{\gamma}_{0,t}$ corresponds to a market portfolio. When each characteristic is also cross-sectionally standardized to have zero mean, this portfolio becomes equivalent to an equal-weighted long-only market portfolio. The remaining elements of $\hat{\gamma}_t$ represent slope coefficients on the characteristics. For example, $\gamma_{BM,t}$ is the coefficient on the book-to-market ratio, $Z_{BM,t}$. These slope coefficients capture the returns on long-short, zero-investment portfolios (cross-sectional factors) when Z_t includes an intercept or is standardized to have a mean of zero.

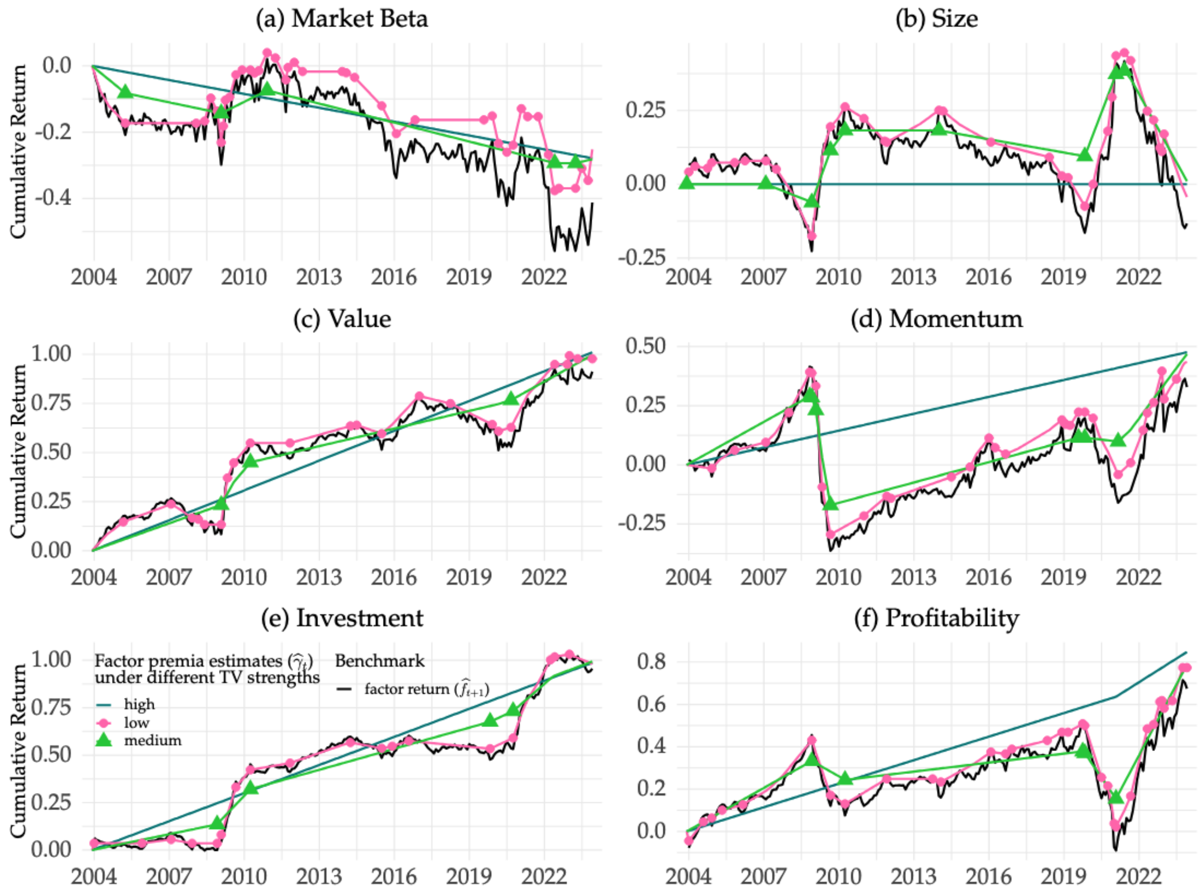
Identifying the underlying trends in these factor premia requires navigating a crucial trade-off. If coefficients change excessively, such as in momentum, the estimator may be overwhelmed by noise introduced by frequent structural breaks. Conversely, prolonged periods of constant coefficients may obscure meaningful structural changes, leading to missed detections. Our approach proxies this by different regularization hyperparameters. In this section, we use results across different hyperparameters to demonstrate that both structural breaks and trend dynamics are present, and then conduct the investment exercises to evaluate which breaks and trends are economically meaningful.[28] In the next section, we implement a machine–learning–based tuning procedure within a multivariate framework and demonstrate that it yields substantial gains in predictive accuracy and robustness compared to univariate approaches. The optimal change frequency strikes a balance between the ability to detect structural breaks and the need to construct a parsimonious, piecewise-constant representation of the underlying trend.

Our analysis focuses on six standard asset pricing factors, including Market (measured by CAPM beta), Size (proxied by market equity), Value (captured through the book-to-market ratio), Momentum (evaluated over a 12-month horizon), Investment (defined as the asset growth rate), and Profitability (represented by operating profitability), to study the dynamics of their risk premia.

[28]Univariate regressions struggle to predict individual stock returns due to low signal-to-noise ratios. We evaluate performance across varying tuning parameters, aiming to provide a comprehensive analysis that illustrates the breaks and trends.

Figure 4: Cumulative Return of Time-varying Risk Premium and Slope Factors

Note: This figure presents cumulative returns for six slope factors and their corresponding time-varying factor premia estimates, evaluated under varying regularization levels. Structural shifts in the data are highlighted through estimated breakpoints. Eq. (1) is estimated using a sample spanning the period from 2004 to 2023, a time marked by significant economic and financial changes.

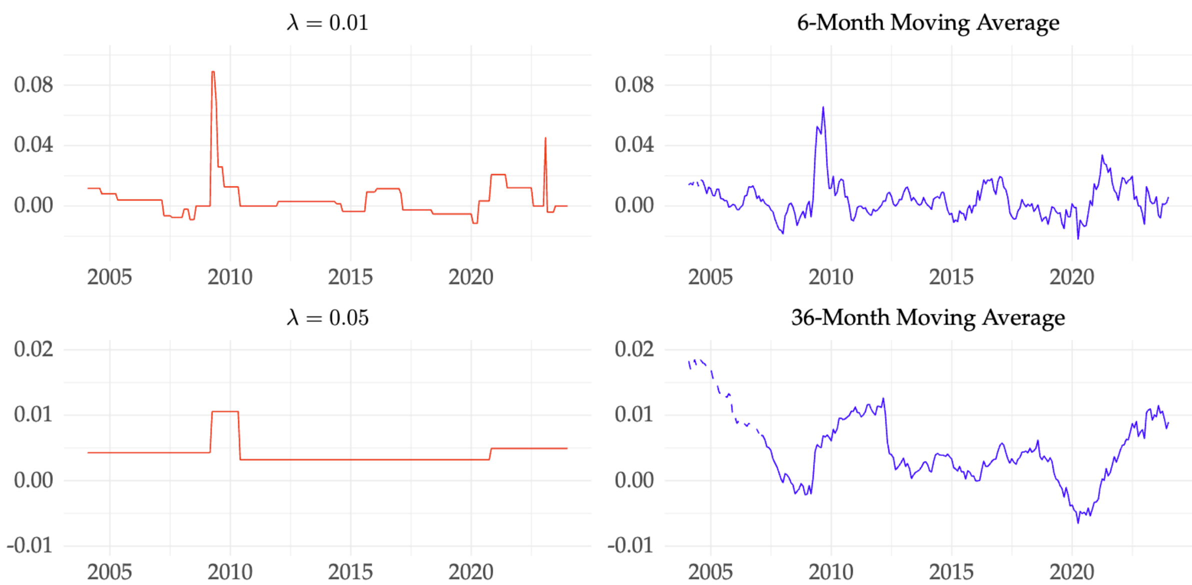


In Figure 4, factor premia estimated with medium regularization (green line) effectively capture regime shifts, such as the global financial crisis and the COVID-19 shock, while also reflecting trends in realized factor returns. This suggests that TV-PR identifies regime-dependent factor premia. Meanwhile, low regularization (pink line) yields too many breaks, causing it to track the factor estimation ($\hat{f}_{t+1}$, black line) rather than the underlying risk premium. In contrast, high regularization (blue line) yields nearly constant estimates that capture only broad trends while ignoring structural breaks. In addition to common break points associated with major economic events, individual factor premia exhibit distinct patterns. For instance, the size factor shows a unique structural break at the beginning of 2014, which is absent in other factors under medium regularization. Additionally, Panel (d) of Figure 4 highlights momentum crash periods, marked by sharp declines in profitability during market reversals. These findings demonstrate TV-PR's ability to capture both broad market disruptions and factor-specific dynamics.

We next zoom in on the one-factor premium to show piecewise-constant patterns and structural break points, and to further assess its sensitivity to the level of regularization. Compared to model-free approaches, such as simple moving averages of cross-sectional factors, our piecewise-constant framework offers greater precision in detecting structural breaks associated with specific events.

Figure 5: Time-Varying Risk Premium Estimation for BE/ME

Note: This figure illustrates the time-varying risk premium of the value factor, proxied by the book-to-market ratio. We compare four approaches: two derived via our methodology under varying regularization levels, and the 6-month and 36-month moving averages. Dotted lines denote the initial estimation sample used for computing the moving averages.



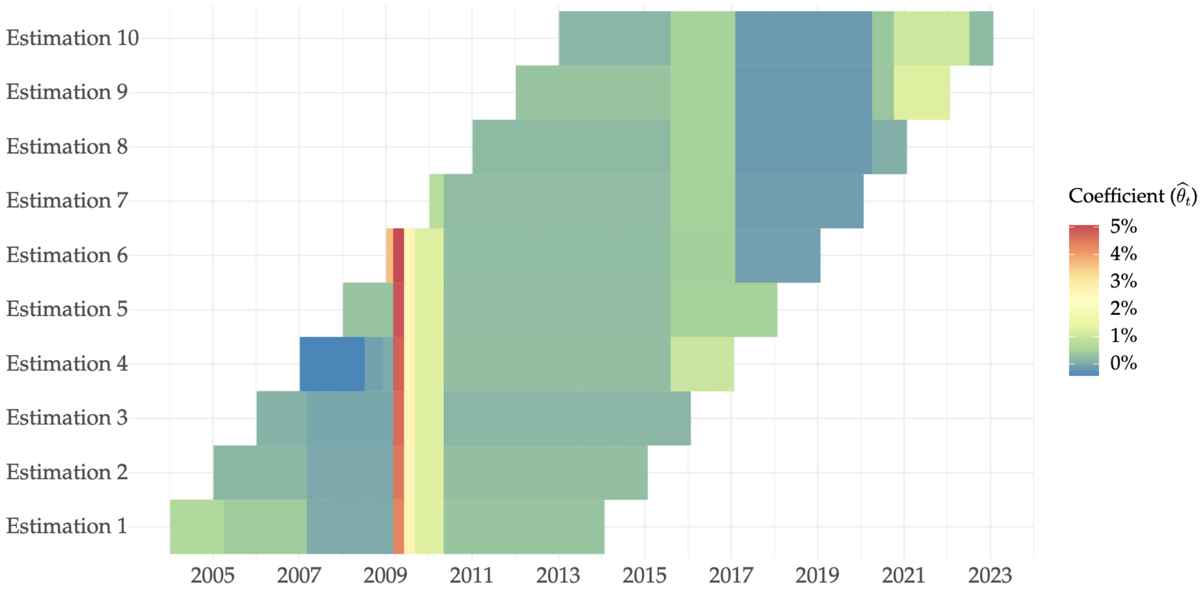
The low regularization level in our method resembles the behavior of a short-window moving average, but with distinct structural break points. As shown in the upper panel of Figure 5, our estimation ($\lambda = 0.01$) aligns closely with a 6-month moving average. However, there are two important differences between the TV-PR estimates and the moving average. First, the piecewise-constant part in factor premia clearly captures the trend shown in Figure 4 when viewed through cumulative returns. In contrast, the moving-average estimates continuously evolve, capturing both factor premia and factor innovations, and thus lead to a factor–return rather than a factor–premia estimation. Second, TV-PR highlights structural changes through abrupt breaks, rather than gradual changes, helping us identify changes in regimes. These differences are more salient at a medium regularization level ($\lambda = 0.05$), where TV-PR captures major peaks and dips, such as those around 2010 and 2020, as illustrated in the left lower panel of Figure 5. On the contrary, although a 36-month moving average reflects these changes, TV-PR introduces minor fluctuations that obscure clear breakpoints. Overall, TV-PR provides a more interpretable representation of time-series changes by emphasizing sparse structural shifts. By avoiding overly smooth transitions, it facilitates sharper insights into underlying trends and breaks, offering a valuable alternative to traditional smoothing techniques.

We then analyze the robustness of structural break identification across different estimation windows. If structural breaks reflect transitions in economic fundamentals, such as shifts in inflation dynamics or monetary policy regimes, estimations conducted around these points should consistently detect similar break patterns. We conduct a rolling-window estimation exercise using a ten-year training sample and a one-year testing period. In total, we have ten estimations within the full sample periods.

The rolling-window estimation consistently identifies structural break points, as shown in Figure 6. Across Estimations 1 to 4, a significant break is consistently detected around 2010, accompanied by an increase in the estimated risk premium, highlighted in red. Notably, this break remains identifiable even when the sample period begins during the event, as shown in Estimations 5 and 6. This indicates that TV-PR can detect the break point with relatively few observations around the event. Subsequent estimations (Estimation 6 through Estimation 10) capture additional breakpoints, including one near 2017 and another around 2020, as shown in deep green. Although Estimation 5 does not detect the one in 2017, the later estimations identify it quickly. This suggests that TV-PR delivers real-time break detection with minimal delay, similar to the findings in Smith and Timmermann (2021), and can therefore pro-

Figure 6: Stability of Break Identification for BE/ME

Note: This figure presents 10-year rolling estimates of the dynamic risk premium for the value factor, proxied by the book-to-market ratio. To ensure time-series independence, estimates are computed with a one-year gap between windows. The regularization parameter is set to $\lambda = 0.05$, balancing model flexibility and control over overfitting.



vide timely guidance for investment decisions. Moreover, the estimated risk premia within each regime remain consistent across different sample periods. For instance, risk premia estimates between 2011 and 2015 consistently approximate 0.5% across all ten estimations. These findings highlight the reliability of TV-PR in identifying shifts in risk premia with different sample periods.

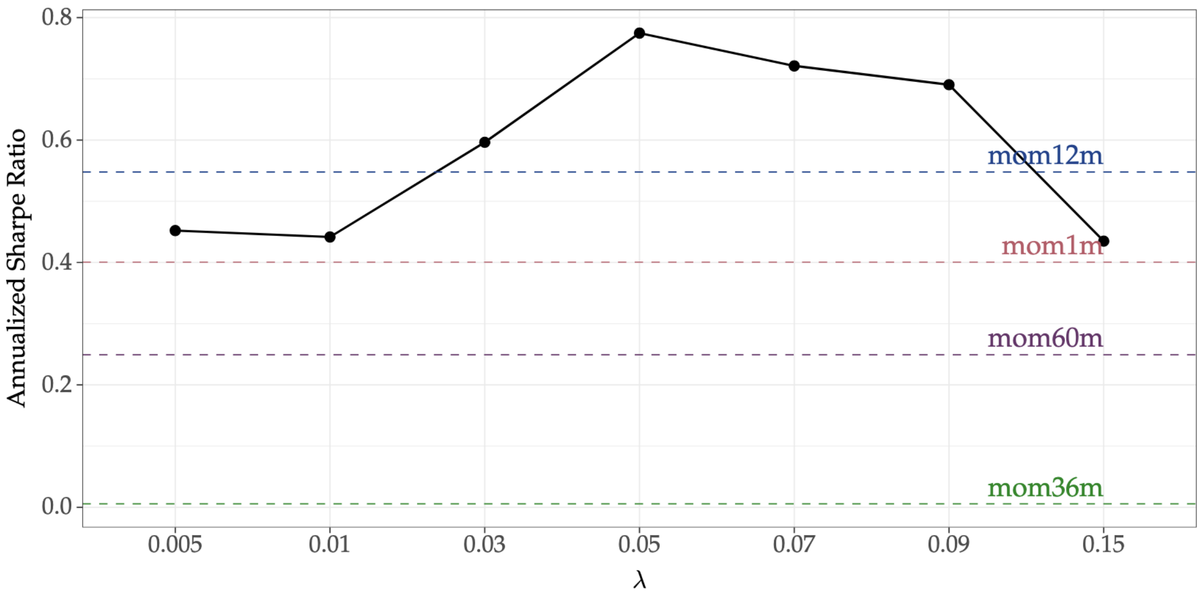
4.3 Factor Timing Investment Performance

Previous results suggest that the estimated factor premia exhibit meaningful economic patterns in both sign and magnitude. The piecewise-constant nature of our estimates provides a framework for predicting out-of-sample risk premia under the assumption that γ_t and γ_{t+1} remain unchanged within the regime. To evaluate their investment performance, we implement a *Factor Sorting* strategy, which ranks factors based on the sign and magnitude of the estimated premia. This strategy assumes that regime-dependent momentum trends in factor premia persist over time. By detecting structural changes with minimal delay, our approach seeks to capture profitability associated with these trends.

In Eq. (10), the left-hand side (LHS) is rearranged to define the cross-sectional factor, where γ_t represents the time-varying risk premia. Based on the time-varying risk premia estimates, we construct a factor-sorting strategy by ranking the monthly values of $\hat{\gamma}_t$. This strategy takes long positions in the five cross-sectional factors with the highest estimated premia and short positions in the five with the lowest. We use a ten-year rolling window of historical data and applies the latest estimate of γ_t to forecast next-month returns. Estimations are updated monthly, with OOS testing sample covering 120 months (2014-2023).

Figure 7: Factor Sorting Strategy Performance based on Time-Varying Premia

Note: This figure reports annualized Sharpe ratios for out-of-sample factor rotation strategies. We construct long-short portfolios using 61 slope factors, with time-varying premia estimated and portfolios rebalanced monthly. Strategy performance is analyzed under varying levels of fused penalty regularization and benchmarked against long-short portfolios based on factor momentum constructed from cumulative past factor returns.



The performance of factor sorting strategies also depends critically on the frequency of parameter estimation. We present the investment performance at various regularization levels. In Figure 7, the optimal parameter achieves the highest Sharpe ratio of 0.78, demonstrating a balance between flexibility and performance. More importantly, the investment performance confirms the economic meaning of the identified breaks and trends in factor premia. The Sharpe ratio remains relatively stable across the hyperparameter spectrum, exhibiting a hump-shaped pattern, indicating that the strategy is robust to different smoothing choices.

The horizontal lines in the figure represent the annualized Sharpe ratios of factor momentum strategies constructed using the same bucket sizes as previously described, where the sorting signal is the past cumulative return of the factor. None of these strategies surpasses TV-PR, reassuring the importance of piecewise-constant patterns in estimating risk premia. Momentum strategies, by contrast, depend on an ad hoc selection of window length, which lacks guidance for optimal specification. TV-PR, on the other hand, identifies piecewise-constant patterns in risk premia and consistently outperforms both over-smoothed (for example, 1-month factor momentum) and under-smoothed (for example, 60-month factor momentum) alternatives, demonstrating strong robustness and adaptability.

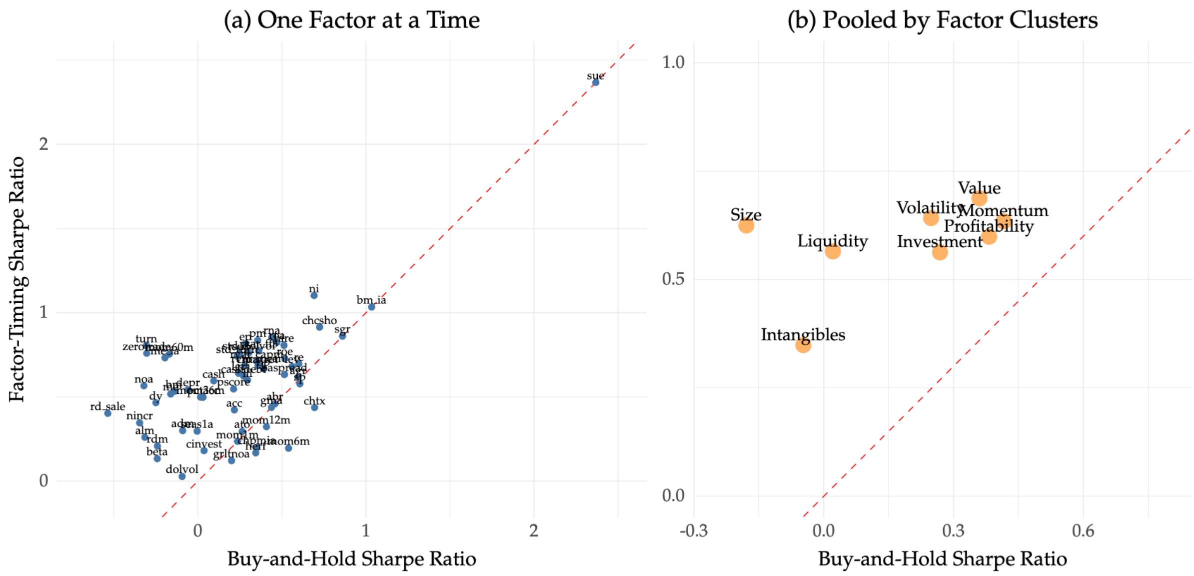
Beyond the factor sorting strategy, Eq. (10) also provides a basis for a timing strategy: invest in the slope factor based on the sign of the estimated $\hat{\gamma}_t$. Previous evidence suggests that these estimates can detect structural breaks and capture trends in factor returns, making them valuable for generating trading signals.

To implement the *Factor Timing* strategy, we take a long position in the cross-sectional factor ($\hat{f}_{t+1}$) when the estimated time-varying risk premium for a given characteristic is positive ($\hat{\gamma}_t > 0$), and a short position when it is negative. As a benchmark, we calculate the buy-and-hold cross-sectional factor return as $(Z_t'Z_t)^{-1}Z_t'r_{t+1} \times$ sgn(chars), where sgn(chars) equals 1 or -1, depending on the direction specified in the literature for each characteristic.

The timing strategy demonstrates superior performance compared to the buy-and-hold benchmark across most characteristics, as indicated by the majority of points lying above the 45-degree line in Panel (a) of Figure 8. Panel (b) categorizes the slope factors into eight groups, revealing that firms within the size category achieve the highest returns under the factor-timing strategy. The improvements in Sharpe ratios vary across characteristics, reflecting the heterogeneity in factor premia. Liquidity-related characteristics, such as share turnover (`turn`) and the number of zero-trading days (3-month rolling, `zerotrade`), exhibit the largest gains. This finding highlights

Figure 8: Single Factor Timing Performance on Time-Varying Premia

Note: This figure displays annualized Sharpe ratios for out-of-sample factor timing strategies versus buy-and-hold across 61 slope factors. Timing strategies, rebalanced monthly, adjust exposures based on the sign of estimated time-varying risk premia. For comparison, we report average Sharpe ratios across eight factor categories.



that cross-sectional factors tied to liquidity benefit substantially from timing strategies. Only a small subset of characteristics shows performance that is similar to or slightly below the buy-and-hold benchmark. Overall, the results show that the TV-PR risk-premia estimates have clear economic meaning and confirm that the identified trends and breaks reflect genuine shifts in the underlying factor-premia process.

5 TV-PR with Multiple Characteristics

5.1 Time-Varying Factor Models

One important question in empirical asset pricing is determining when and which factors are useful for pricing other assets, or, equivalently, when these factors are relevant to asset-pricing factor models. To address this question, we analyze the dynamic structure of slope-factor risk premia and compare time-varying and time-invariant factor models.

We utilize 61 firm characteristics to estimate the time-varying risk premia in Eq. (1). To account for both time variation and variable selection during estimation, we employ a combination of total variation and lasso regularizations, as specified in Eq. (6). These penalties strike a balance between sparsity and smoothness in the estimated coefficients, ensuring robust predictions. The hyperparameters associated with the penalties are optimized through OOS cross-validation to maximize predictive accuracy, measured by OOS predictive R^2. Using a rolling window approach, we train the model on a 120-month period and forecast the subsequent 12 months. This method captures recent dynamics while maintaining predictive reliability.

Table 3: Stock Return Predictability

Note: This group is categorized into one of five categories: Mega, Large, Small, Micro, and Nano Cap. The groups are non-overlapping, and the breakpoints are based on the market equity of NYSE stocks by the end of each month. In particular, mega caps are stocks with a market cap above the 80th percentile of NYSE stocks; large caps are all remaining stocks above the 50th percentile; small caps are above the 20th percentile; micro caps are above the 1st percentile; and nano caps are the remaining stocks.

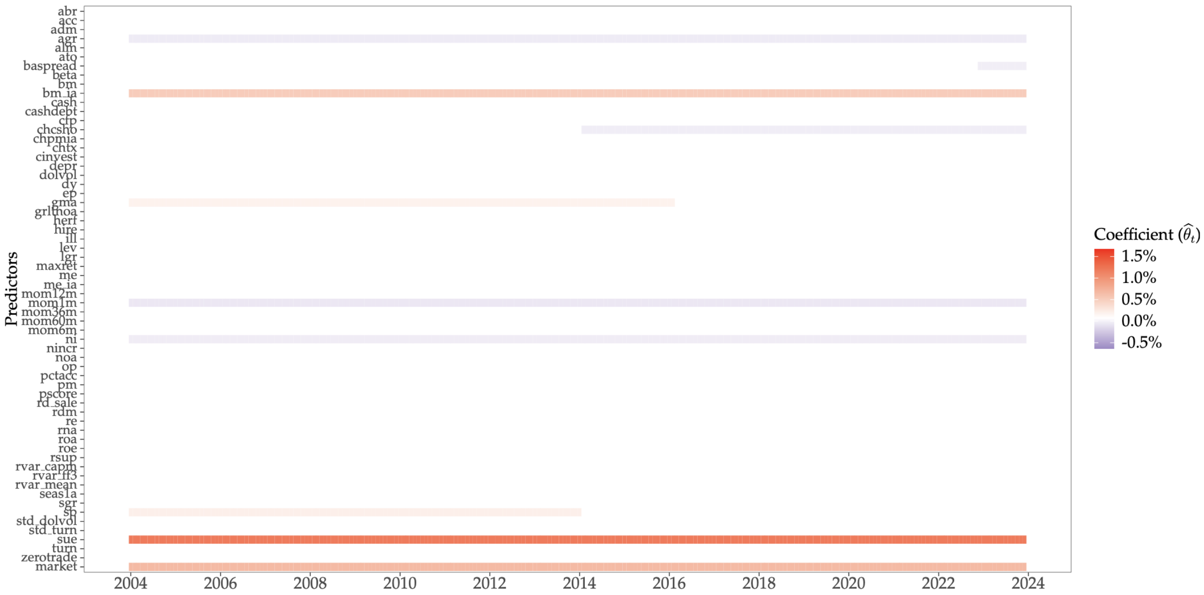
Size group	Mega	Large	Small	Micro	Nano	All
Predictive R^2 (OOS, %)	0.73	0.76	0.51	0.42	-1.05	0.14

Table 3 reports the stock return predictability across different stock-size groups with the pair of hyperparameters yielding a positive overall OOS predictive R^2. The results show that mega and large firms exhibit relatively high OOS R^2, whereas the predictive performance declines for small and micro stocks. The nano group generates a negative OOS R^2, indicating that extremely small firms cannot be reliably predicted. Overall, TV-PR demonstrates meaningful predictability for individual stock returns.

Next, we focus on the full sample analysis on the time-varying risk premia to explore the persistence and variability of each characteristic's impact on expected returns. In Figure 9, among the characteristics analyzed, standardized unexpected earnings (`sue`) emerge as a robust predictor of cross-sectional stock returns, with a monthly risk premium exceeding 1.1%. This finding suggests the importance of earnings surprises in explaining equity returns. The industry-adjusted book-to-market ratio (`bm_ia`) also contributes positively, emphasizing the relevance of sector-neutral valuation measures in asset pricing. Other characteristics, such as sales-to-price (`sp`), gross profitability (`gma`), and change in shares outstanding (`chcsho`), exhibit positive

Figure 9: Time-Varying Factor Premia

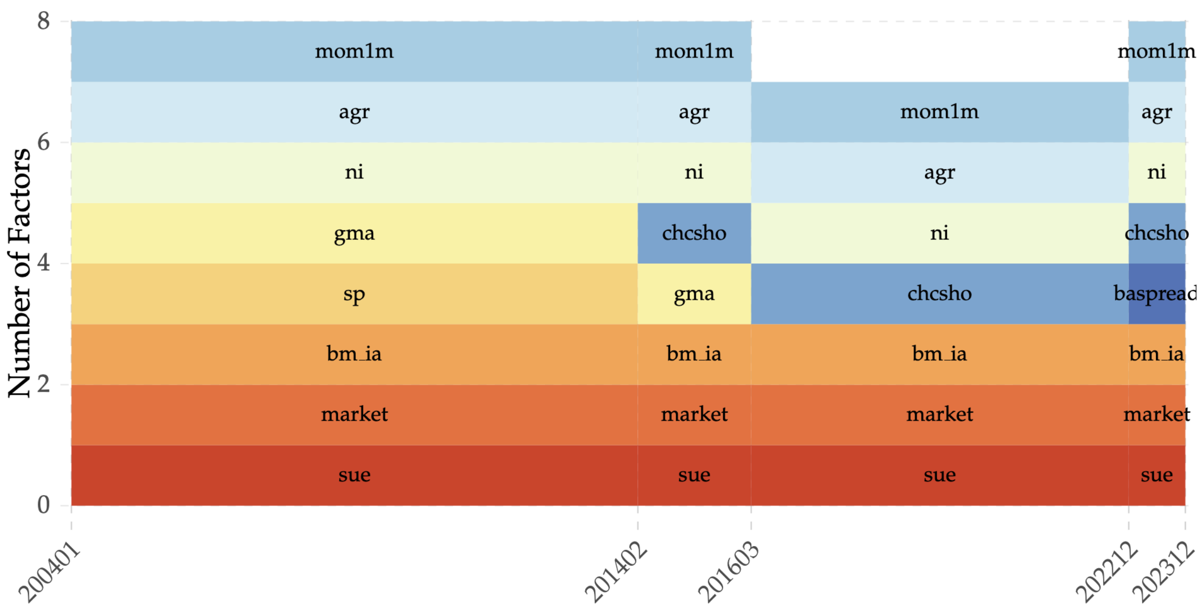
Note: This figure depicts time-varying premia linked to slope factors, estimated from the dynamic coefficients in Eq. (1), which map factor returns to firm-specific attributes. The color bar illustrates the magnitude and thresholds of key firm characteristics, highlighting their relative importance. The intercept represents the equally weighted market portfolio, serving as the benchmark.



risk premia, although their contributions fluctuate over time.

Figure 10: Regime-Specific Factor Model Specification

Note: The figure illustrates the number of selected factors used in asset pricing models within each regime identified by structural breaks. These factors are ranked by their estimated risk premia in descending order, starting from the bottom of the plot.



Following the results above, TV-PR provides a framework for detecting changes in factor model specifications within different distinct economic regimes ($\hat{\theta} \neq 0$). The specification of the factor model in different regimes is crucial for understanding the relationship between the composition of SDF and asset returns. Figure 10 illustrates the regime-specific factor model specification. The model initially included eight factors in early 2004, maintaining this count through 2014. By early 2016, the number of factors had decreased slightly to seven, remained stable through late 2022, and then returned to eight.

The composition of pricing factors evolves significantly over time, even as their overall number remains relatively stable. A core set of factors, including `sue`, `market`, `bm_ia`, and `mom1m`, consistently demonstrates pricing power throughout the two-decade span. In contrast, other factors exhibit transience. For example, `sp`, which was relevant in 2004, is replaced by `chcsho` starting in 2014. The emergence of the `baspread` factor in late 2022 further highlights the dynamic nature of firm characteristics. The results suggest that the attributes influencing asset pricing change over time, reflecting shifts in market conditions and investor preferences.

To evaluate the relative performance of time-varying versus time-invariant factor model specifications, we employ differences in squared Sharpe ratios (SR^2), following the non-nested model comparison framework proposed by Barillas et al. (2020). This methodology enables a rigorous evaluation of how effectively each specification captures evolving market dynamics. The findings underscore the effectiveness of time-varying models in adapting to shifts in market conditions.

The performance of factor models varies across distinct market regimes. As shown in Table 4, during the first regime (2004:01–2014:01), the time-varying specification delivers both statistically and economically superior performance relative to benchmark models. Specifically, the results show substantial SR^2 improvements, such as a 1.48 increase over the CAPM, along with similarly pronounced enhancements relative to FF3, FF5, and other models. These findings are supported by p-values effectively equal to zero, highlighting their statistical significance.

Table 4: Regime Specific Factor Model Comparison

Note: This table presents squared differences in Sharpe ratios between time-varying and time-invariant factor models across market regimes. Time-invariant models include CAPM, Fama-French 3-factor (FF3, Fama and French, 1993), 5-factor (FF5, Fama, 2015), 6-factor (FF6, Fama and French, 2018), Hou et al. (2021) 5-factor ($q5$), Daniel et al. (2020) 3-factor (DHS3), and Barillas and Shanken (2018) 6-factor (BS6) model. p-values are reported in parentheses.

	0401-1401	1402-1602	1603-2211	2212-2312
CAPM	1.48	-0.10	-0.09	-0.18
	(0.00)	(0.54)	(0.17)	(0.40)
FF3	1.49	-0.07	-0.07	-0.05
	(0.00)	(0.68)	(0.29)	(0.84)
FF5	1.41	-0.13	-0.08	0.04
	(0.00)	(0.60)	(0.30)	(0.87)
FF6	1.42	-0.15	-0.08	0.13
	(0.00)	(0.51)	(0.33)	(0.57)
$q5$	1.40	-0.55	-0.12	-0.04
	(0.00)	(0.07)	(0.21)	(0.88)
DHS3	1.45	-0.26	-0.10	-0.20
	(0.00)	(0.26)	(0.21)	(0.56)
BS6	1.45	-0.23	-0.08	0.12
	(0.00)	(0.32)	(0.34)	(0.59)

In contrast, the incremental value added by the time-varying model diminishes considerably in post-2014 regimes. Across most specifications and subperiods from 2014 onward, SR^2 differences are economically modest and statistically insignificant. Furthermore, distinctions among time-invariant models become less pronounced in later regimes, highlighting the increased difficulty in differentiating between competing factor models during these periods. This observation applies not only to comparisons between time-varying and time-invariant models but also to distinctions among the time-invariant models themselves.

5.2 Forecast-Implied Portfolio Investment Strategy

Although academic research has identified numerous return predictors with in-sample predictive power, their OOS performance often lacks consistency, raising concerns about their practical applicability. This instability likely stems from fundamental shifts in economic regimes and structural changes in risk-return trade-offs rather than mere statistical artifacts. Balancing static and time-varying, unconditional and conditional models is essential to avoid model misspecification and capture the dynamic nature of expected returns. This study investigates whether explicitly modeling time variation can yield consistent and economically significant OOS investment performance.

To examine this question, we construct investment portfolios based on predicted returns ($\hat{r}_{i,t+1}$) as signals for portfolio allocation in subsequent periods. The model is re-estimated annually using a rolling window of the past ten years. By explicitly accounting for structural shifts in market dynamics, our methodology adapts to evolving conditions, offering a practical framework for improving asset allocation decisions. Meanwhile, we employ period-by-period regression models as a benchmark to capture maximum flexibility in our analysis.

Forecast-implied portfolios take long positions in stocks in the top decile of predicted returns and short positions in stocks in the bottom decile. Both equal-weighted and value-weighted approaches are employed to allocate positions.

Table 5: Out-of-Sample Forecast-Implied Investment Performance

Note: This table presents the OOS performance of forecast-implied portfolios, constructed using equal-weighted and value-weighted long-short strategies. The analysis employs rolling regressions, including rolling OLS, rolling LASSO, and TV-PR methods. Each panel reports five key performance metrics: monthly average return (Avg, %), standard deviation (Std, %), annualized Sharpe ratio (SR), Jensen's alpha (%), and maximum drawdown (MDD, %).

	Avg	Std	SR	Alpha	MDD
	Panel A: Equal-Weighted				
Rolling OLS	2.61	3.87	2.34	2.87***	16.54
Rolling LASSO	2.04	3.78	1.87	2.31***	17.67
TV-PR	2.46	3.40	2.50	2.60***	6.41
	Panel B: Value-Weighted				
Rolling OLS	0.22	4.64	0.16	0.57	52.06
Rolling LASSO	0.05	4.19	0.04	0.44	47.75
TV-PR	0.43	3.41	0.44	0.70**	29.06

Table 5 summarizes the OOS performance of forecast-implied portfolio strategies. In Panel A, the equal-weighted long-short portfolio based on TV-PR demonstrates strong performance, with an annualized Sharpe ratio of 2.5 and a monthly alpha of approximately 2.6%. TV-PR also exhibits a significantly lower maximum drawdown compared to the period-by-period regression approach. Similarly, Panel B shows that the value-weighted long-short portfolio constructed using TV-PR achieves a higher

Sharpe ratio and a statistically significant alpha, while maintaining a considerably smaller maximum drawdown than the period-by-period regression method.[29] These results underscore the model's predictive accuracy and economic value.

The forecast-implied performance demonstrates that introducing structural breaks at appropriate frequencies enhances the model's ability to capture dynamic changes. While period-by-period regression updates effectively adapt to frequent structural changes, they result in increased variance, as evidenced by a lower Sharpe ratio and significant maximum drawdown. By comparison, the TV-PR model achieves an optimal balance in the bias-variance trade-off while maintaining an interpretable factor structure that enhances investment performance.

5.3 Return-Based Anomalies

Empirical asset-pricing studies often document cross-sectional return predictability using a fixed set of return-based predictors, implicitly assuming a time-invariant relationship between expected return premia and these predictors (e.g., Fama and French, 2008; Novy-Marx, 2013). In this section, we re-examine the dynamic nature of these relationships by studying the evolution of risk premia associated with return-based predictors over time. Return-based predictors are particularly suited for this analysis due to their well-defined nature, long history, and systematic applicability both before and after publication. Specifically, we analyze lagged returns over short-term horizons (one to twelve months) to capture reversal and momentum effects. To detect seasonality patterns, we examine lagged returns at intervals of 24, 36, 48, and 60 months. Additionally, cumulative returns over months 13–60 are analyzed to identify long-term reversal effects.[30]

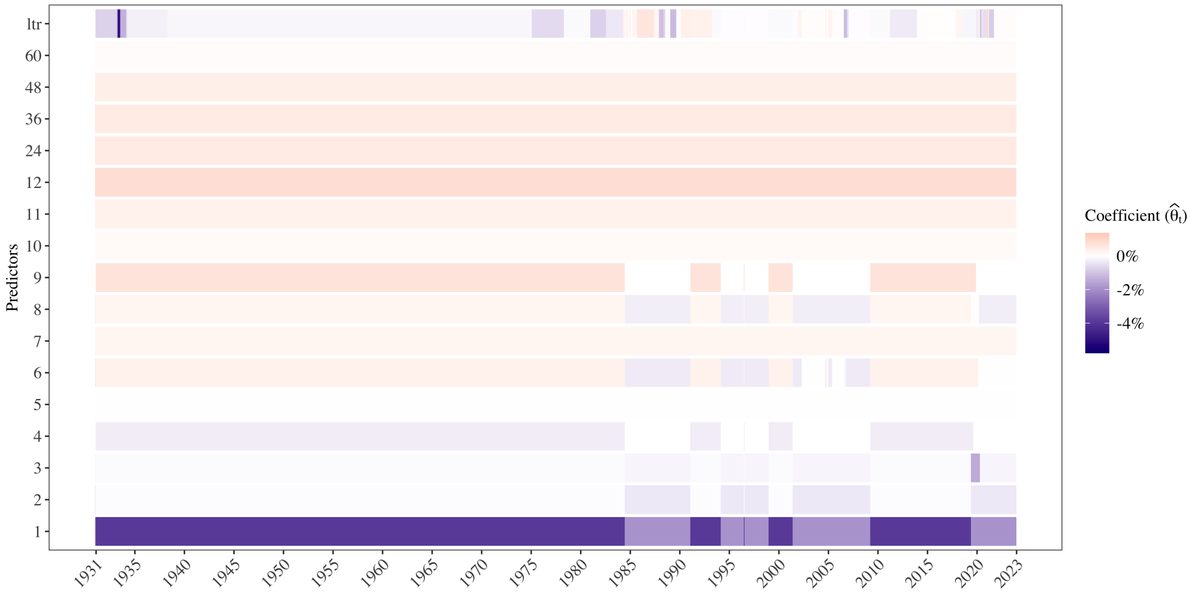
Figure 11 plots the estimated time-varying coefficients of return-based predictors from 1931 to 2023. The long-term reversal effect, as documented in De Bondt and Thaler (1985), is prominently visible in the `ltr` predictor during the early-20th cen-

[29] The greater maximum drawdown experienced by the value-weighted portfolio, compared to its equal-weighted counterpart, is largely driven by substantial losses incurred during the COVID-19 pandemic in 2020.

[30] Following Marrow and Nagel (2024), we exclude firms with market capitalization below the 20th percentile of the prior month-end.

Figure 11: Time-Varying Coefficients of Return-Based Predictors

Note: This figure illustrates dynamic premia associated with return-based anomalies, estimated using dynamic coefficients from Eq. (1). The predictors include 1-12, 24, 36, 48, and 60 lagged returns. `ltr` is the long-term reversal effects with cumulative returns over months 13-60. The color bar indicates the magnitude and thresholds of these predictors, highlighting their significance in explaining return anomalies.



tury but weakens substantially over time. By the late 20th century, this effect became volatile and exhibited occasional opposite signs. Similarly, the 2–12-month momentum premium identified by Jegadeesh and Titman (1993) undergoes notable evolution. Its profile shifts structurally in subsequent decades, with the effect concentrating in the 7–12-month window and attenuating at shorter lags, consistent with the findings in Novy-Marx (2012). In contrast, the one-month reversal documented by Jegadeesh (1990) demonstrates remarkable persistence throughout the sample period. Despite minor variations in strength, its negative coefficient remains consistently strong across decades. The results indicate that the seasonality effect at the 24-, 36-, 48-, and 60-month lags, as identified by Heston and Sadka (2008), exhibits a consistent pattern with no changes. Overall, the results emphasize that while anomalies may appear ex ante before their publication, their predictive power often diminishes following academic discovery.

From a technical trading perspective, time-varying factor premia are reflected by observable trends in factor returns. TV-PR formalizes this concept by incorporating investors' prior beliefs about the smoothness of these trends and the frequency of market regime changes. The framework employs total-variation regularization to represent these prior beliefs. A higher penalty reflects an assumption of smooth, long-term trends, producing piecewise-linear trend lines that highlight stable market regimes with major structural breaks—similar to technical traders using long-term moving averages. Conversely, a lower penalty captures more volatile trends, enabling the model to identify subtle, short-term shifts in market dynamics.

This approach translates the varying time horizons and trend-smoothness assumptions inherent in technical analysis into a disciplined statistical model. Additionally, by applying a second Lasso penalty for factor selection, the framework accounts for the transient relevance of factors, allowing predictive signals to adapt dynamically. This provides a robust method for systematically identifying regime-dependent momentum trends and structural breaks based on specified investor beliefs about market dynamics. TV-PR offers an alternative to Bayesian shrinkage methods, such as those proposed by Marrow and Nagel (2024), by integrating investor assumptions directly into the model's structure.

6 Conclusion

This paper develops a regime-based predictive regression framework that accommodates both structural breaks and time-varying sparsity in the relationship between firm characteristics and expected returns. By combining cross-sectional information with total variation and Lasso penalties, the proposed TV-PR estimator delivers piecewise constant risk premia and dynamically selects predictors without relying on restrictive assumptions about coefficient evolution. This design allows the model to adjust flexibly to changes in underlying economic conditions while maintaining parsimony.

Our theoretical results establish the large-sample properties of the estimator under settings with both known and unknown structural breaks, providing a foundation for inference on dynamic relationships. Empirically, the framework reveals economically meaningful shifts in factor premia, underscoring the transient nature of return predictors. A factor-timing strategy based on TV-PR's estimated premia generates higher performance than buy-and-hold and factor momentum benchmarks. These findings show that accounting for breaks and trend dynamics improves real-time predictive accuracy.

Compared to static models, TV-PR offers superior explanatory power and improved investment performance, particularly during periods of structural change. By incorporating regime-dependent dynamics and time-varying sparsity, TV-PR addresses key complexities in return prediction. These findings contribute to the literature on dynamic asset pricing models and offer practical insights for portfolio management.

References

An, Y., Y. Su, and C. Wang (2025). Quantity, risk, and return. Technical report, Johns Hopkins University.

Arnott, R. D., V. Kalesnik, and J. T. Linnainmaa (2023). Factor momentum. *Review of Financial Studies 36*(8), 3034–3070.

Avramov, D. (2004). Stock return predictability and asset pricing models. *Review of Financial Studies 17*(3), 699–738.

Bandi, F. M. and Y. Su (2025). Conditional spectral methods. *Journal of Econometrics 248*, 105863.

Bandi, F. M. and A. Tamoni (2022). Spectral financial econometrics. *Econometric Theory 38*(6), 1175–1220.

Barillas, F., R. Kan, C. Robotti, and J. Shanken (2020). Model comparison with sharpe ratios. *Journal of Financial and Quantitative Analysis 55*(6), 1840–1874.

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

Barroso, P., M. Boons, and P. Karehnke (2021). Time-varying state variable risk premia in the icapm. *Journal of Financial Economics 139*(2), 428–451.

Bickel, P. J., Y. Ritov, and A. B. Tsybakov (2009). Simultaneous analysis of lasso and dantzig selector. *Annals of Statistics 37*(4), 1705–1732.

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

Bryzgalova, S., M. Pelger, and J. Zhu (2020). Forest through the trees: Building cross-sections of stock returns. Technical report, London Business School.

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

Chen, Q., N. Roussanov, and X. Wang (2025). Semiparametric conditional factor models in asset pricing. Technical report, The Chinese University of Hong Kong, Shenzhen.

Chib, S., Y. C. Lin, K. Pukthuanthong, and X. Zeng (2023). Slope factors outperform: evidence from a large comparative study. Technical report, Washington University in St. Louis.

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

Dai, C., K. Lu, and D. Xiu (2019). Knowing factors or factor loadings, or neither? evaluating estimators of large covariance matrices with noisy and asynchronous data. *Journal of Econometrics 208*(1), 43–79.

Dangl, T. and M. Halling (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics 106*(1), 157–181.

Daniel, K., D. Hirshleifer, and L. Sun (2020). Short-and long-horizon behavioral factors. *Review of Financial Studies 33*(4), 1673–1736.

Daniel, K. and T. J. Moskowitz (2016). Momentum crashes. *Journal of Financial Economics 122*(2), 221–247.

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

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

Ehsani, S. and J. T. Linnainmaa (2022). Factor momentum and the momentum factor. *Journal of Finance 77*(3), 1877–1919.

Fama, E. (1976). Foundations of finance: Portfolio decisions and securities prices.

Fama, E. F. (2015). Cross-section versus time-series tests of asset pricing models. Technical report, University of Chicago.

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 (2008). Dissecting anomalies. *Journal of finance 63*(4), 1653–1678.

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

Fama, E. F. and K. R. French (2020). Comparing cross-section and time-series factor models. *Review of Financial Studies 33*(5), 1891–1926.

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.

Fan, J., Y. Liao, and W. Wang (2016). Projected principal component analysis in factor models. *Annals of statistics 44*(1), 219.

Farmer, L. E., L. Schmidt, and A. Timmermann (2023). Pockets of predictability. *Journal of Finance 78*(3), 1279–1341.

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

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.

Gagliardini, P., F. Trojani, and G. Urga (2005). Robust GMM tests for structural breaks. *Journal of Econometrics 129*(1-2), 139–182.

Giglio, S. and D. Xiu (2021). Asset pricing with omitted factors. *Journal of Political Economy 129*(7), 1947–1990.

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

Harchaoui, Z. and C. Lévy-Leduc (2010). Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association 105*(492), 1480–1493.

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

Henkel, S. J., J. S. Martin, and F. Nardari (2011). Time-varying short-horizon predictability. *Journal of Financial Economics 99*(3), 560–580.

Heston, S. L. and R. Sadka (2008). Seasonality in the cross-section of stock returns. *Journal of Financial Economics 87*(2), 418–445.

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.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance 45*(3), 881–898.

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

Ke, Z. T., J. Fan, and Y. Wu (2015). Homogeneity pursuit. *Journal of the American Statistical Association 110*(509), 175–194.

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

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.

Kock, A. B. and H. Tang (2019). Uniform inference in high-dimensional dynamic panel data models with approximately sparse fixed effects. *Econometric Theory 35*(2), 295–359.

Kozak, S. and S. Nagel (2023). When do cross-sectional asset pricing factors span the stochastic discount factor? Technical report, National Bureau of Economic Research.

Li, D., J. Qian, and L. Su (2016). Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association 111*(516), 1804–1819.

Li, S. Z., P. Yuan, and G. Zhou (2023). Pockets of factor pricing. Technical report, Rutgers Business School.

Marrow, B. and S. Nagel (2024). Real-time discovery and tracking of return-based anomalies. Technical report, National Bureau of Economic Research.

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

Neely, C. J., D. E. Rapach, J. Tu, and G. Zhou (2014). Forecasting the equity risk premium: the role of technical indicators. *Management science 60*(7), 1772–1791.

Novy-Marx, R. (2012). Is momentum really momentum? *Journal of Financial Economics 103*(3), 429–453.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics 108*(1), 1–28.

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

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

Pettenuzzo, D. and A. Timmermann (2011). Predictability of stock returns and asset allocation under structural breaks. *Journal of Econometrics 164*(1), 60–78.

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

Rosenberg, B. (1974). Extra-market components of covariance in security returns. *Journal of Financial and Quantitative Analysis 9*(2), 263–274.

Rudin, L. I., S. Osher, and E. Fatemi (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena 60*(1-4), 259–268.

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.

Smith, S. C., A. Timmermann, and Y. Zhu (2019). Variable selection in panel models with breaks. *Journal of Econometrics 212*(1), 323–344.

Su, L. and X. Wang (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics 198*(1), 84–101.

Tang, X., X. Tang, and G. Zhou (2022). Heterogeneous responses in financial markets: Insights from machine learning. Technical report, University of Texas at Dallas.

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.

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

Yousuf, K. and S. Ng (2021). Boosting high dimensional predictive regressions with time varying parameters. *Journal of Econometrics 224*(1), 60–87.

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

Zhu, Y. and G. Zhou (2009). Technical analysis: An asset allocation perspective on the use of moving averages. *Journal of Financial Economics 92*(3), 519–544.

Zou, H. and R. Li (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Annals of Statistics 36*(4), 1509.

Appendices

A Additional Results

Figure A.1: Factor Realizations from One Simulation

Note: This figure illustrates the realizations of the three factors ($f_{t+1}^{[k]} = \gamma_t^{[k]} + \nu_{t+1}^{[k]},\ k = 1, 2, 3$) from a single simulation ($N = 1000$, $T = 240$).

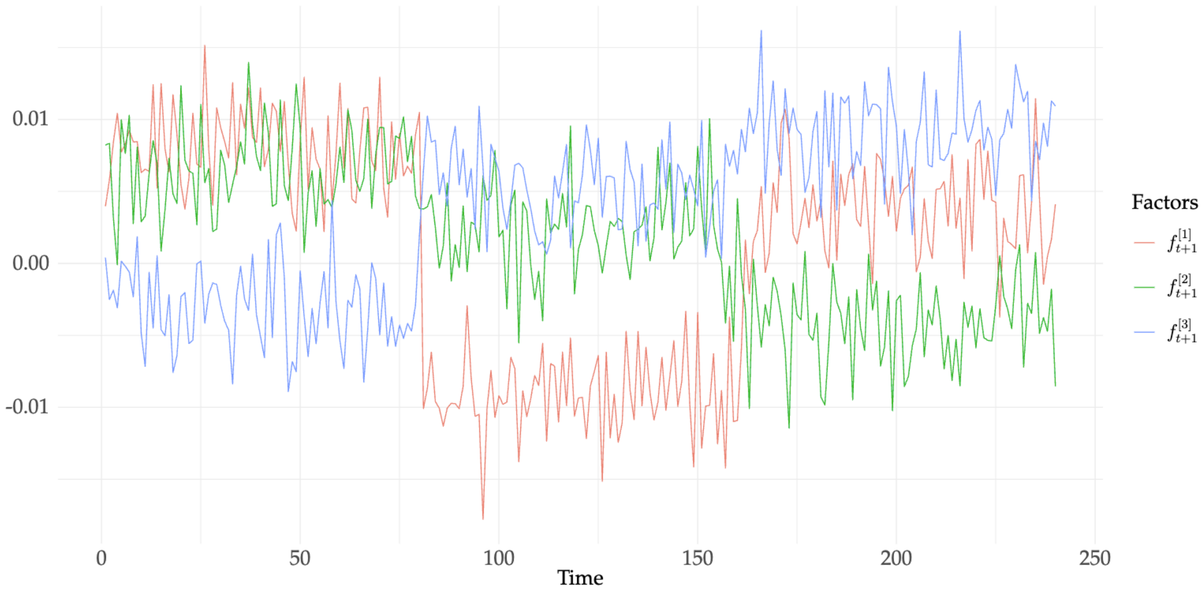


Figure A.1 displays the realizations of all factors across regimes from a single simulation draw. The first factor, $f_{t+1}^{[1]}$, exhibits a decline in expected returns, shifting from positive in the first regime to negative in the second, before returning to positive in the third regime. The second factor, $f_{t+1}^{[2]}$, shows a continuous decline, whereas the third factor, $f_{t+1}^{[3]}$, steadily increases from the first to the final regime. Factor innovations introduce noise into risk premia estimation; thus, to obtain clean estimates, researchers need a methodology to average out these innovations within each regime.

Figure A.2: Sharpe Ratios of Factor-Timing Strategies with Different Hyperparameters

<u>Note:</u> This figure reports annualized Sharpe ratios for the factor-timing strategy versus buy-and-hold strategy using extreme hyperparameter values (largest and smallest) over 2004–2023. The results indicate that the two extremes — frequently adjusting premia every period and maintaining constant premia — underperform the buy-and-hold strategy. This finding suggests that intermediate points between these extremes may deliver superior investment performance, as documented in the main results.

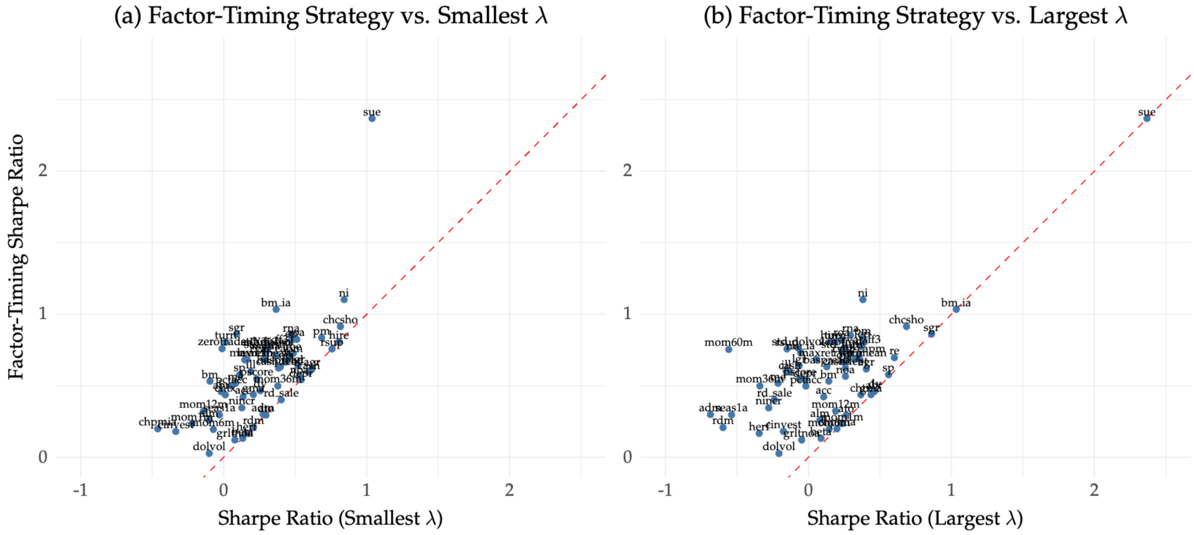
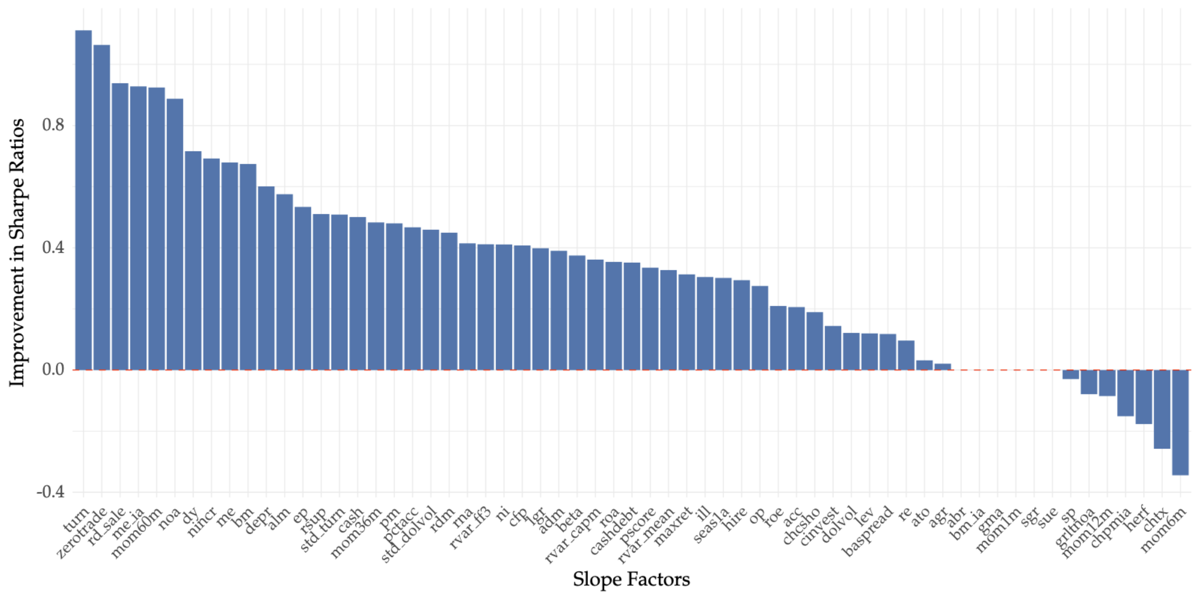


Figure A.3: Investment Improvement of Factor Timing Strategies

<u>Note:</u> This figure reports the improvement in annualized Sharpe ratios of the out-of-sample factor-timing strategy relative to the buy-and-hold benchmark from 2004 to 2023. The factor-timing approach leverages the sign of estimated time-varying risk premia, with monthly portfolio rebalancing.



B Predictor Descriptions

Table A.1: **Equity Characteristics**

No.	Acronym	Description	Category
1	abr	Cumulative abnormal returns around earnings announcement dates	Momentum
2	acc	Operating Accruals	Investment
3	adm	Advertising Expense-to-market	Intangibles
4	agr	Asset growth	Investment
5	alm	Quarterly Asset Liquidity	Intangibles
6	ato	Asset Turnover	Profitability
7	baspread	Bid-ask spread rolling 3m	Liquidity
8	beta	Beta rolling 3m	Volatility
9	bm	Book-to-market equity	Value
10	bm_ia	Industry-adjusted book to market	Value
11	cash	Cash holdings	Value
12	cashdebt	Cash flow to debt	Value
13	cfp	Cash flow to price ratio	Value
14	chcsho	Change in shares outstanding	Investment
15	chpmia	Industry-adjusted change in profit margin	Profitability
16	chtx	Change in tax expense	Momentum
17	cinvest	Corporate investment	Investment
18	depr	Depreciation / PP&E	Momentum
19	dolvol	Dollar trading volume	Liquidity
20	dy	Dividend yield	Value
21	ep	Earnings-to-price	Value
22	gma	Gross profitability	Profitability
23	grltnoa	Growth in long-term net operating assets	Investment
24	herf	Industry sales concentration	Intangibles
25	hire	Employee growth rate	Intangibles
26	ill	Illiquidity rolling 3m	Liquidity
27	lev	Leverage	Value
28	lgr	Growth in long-term debt	Investment
29	maxret	Maximum daily returns rolling 3m	Volatility
30	me	the market equity	Size
31	me_ia	Industry-adjusted size	Size
32	mom12m	Momentum rolling 12m	Momentum
33	mom1m	Momentum	Momentum
34	mom36m	Momentum rolling 36m	Momentum

Table A.1: Equity Characteristics (Continued)

No.	Acronym	Description	Category
35	mom60m	Momentum rolling 60m	Momentum
36	mom6m	Momentum rolling 6m	Momentum
37	ni	Net Stock Issues	Investment
38	nincr	Number of earnings increases	Momentum
39	noa	(Changes in) Net Operating Assets	Investment
40	op	Operating profitability	Profitability
41	pctacc	Percent operating accruals	Investment
42	pm	profit margin	Profitability
43	pscore	Performance Score	Profitability
44	rd_sale	R&D to sales	Intangibles
45	rdm	R&D Expense-to-market	Intangibles
46	re	Revisions in analysts earnings forecasts	Momentum
47	rna	Quarterly Return on Net Operating Assets	
48	roa	Return on assets	Profitability
49	roe	Return on equity	Profitability
50	rsup	Revenue surprise	Momentum
51	rvar_capm	Residual variance - CAPM rolling 3m	Volatility
52	rvar_ff3	Residual variance - ff3 rolling 3m	Volatility
53	seas1a	Seasonality	Intangibles
54	sgr	Sales growth	Value
55	sp	Sales-to-price	Value
56	std_dolvol	Std. of dollar trading volume rolling 3m	Volatility
57	std_turn	Std. of Share turnover rolling 3m	Volatility
58	sue	Unexpected quarterly earnings	Momentum
59	rvar_mean	return variance rolling 3m	Volatility
60	turn	Shares turnover	Liquidity
61	zerotrade	Number of zero-trading days rolling 3m	Liquidity