



HAL
open science

Decomposing anomalies

Sabri Boubaker, Bo Li, Zhenya Liu, Yifan Zhang

► **To cite this version:**

Sabri Boubaker, Bo Li, Zhenya Liu, Yifan Zhang. Decomposing anomalies. *Economics Letters*, 2021, 202, pp.109835. 10.1016/j.econlet.2021.109835 . hal-03323669

HAL Id: hal-03323669

<https://hal.science/hal-03323669>

Submitted on 24 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Decomposing Anomalies[★]

Sabri Boubaker^{a,b}, Bo Li^c, Zhenya Liu^{d,e,*}, Yifan Zhang^{d,1}

^a*EM Normandie Business School, Métis Lab, France*

^b*International School, Vietnam National University, Hanoi, Vietnam*

^c*Business School, Beijing International Studies University, Beijing, China*

^d*School of Finance, Renmin University of China, Beijing, China*

^e*CERGAM, Aix-Marseille University, Aix-en-Provence, France*

*We would like to thank the editor and referee for their helpful comments. All errors remain our own. This research has been supported by the Fundamental Research Funds of the Central Universities, and the Research Funds of Renmin University of China, grant 21XNH010.

*Correspondence to: School of Finance, Renmin University of China, No.59, Zhongguancun Street, Haidian District, 100872, Beijing, China

Email addresses: sboubaker@em-normandie.fr (Sabri Boubaker), libo@bisu.edu.cn (Bo Li), zhenya.liu@ruc.edu.cn (Zhenya Liu), zhangyifan@ruc.edu.cn (Yifan Zhang)

Decomposing Anomalies^{*}

Abstract

This paper introduces the functional principal component analysis approach for decomposing the panel returns of the anomaly-sorted portfolios. Using the US stock market data covering July 1963-July 2020, our findings indicate that the Fama-French (F-F) market factor can be captured by the first empirical functional principal component in the time-series. For the other F-F anomalies, these being market capitalization (*Size*), book-to-market ratio (*B/M*), profitability (*OP*), investment (*Inv*), and price momentum (*Mom*), the cross-sectional features remain in the monotonicity of the second principal component and in the curvature of the third. Furthermore, a time-varying framework shows two neglected reversals of the F-F anomalies *Inv* and *Size* in the 1970s and the 1980s.

Keywords: Asset Pricing, Anomaly Variable, Factor Model, Functional Principal Component Analysis, Eigenfunction

JEL: C23, C51, G11, G12

^{*}We would like to thank the editor and referee for their helpful comments. All errors remain our own.

1. Introduction

The existing literature documents a wide range of anomalies that are not explained by the Capital Asset Pricing Model (CAPM). These include the book-to-market equity (Fama & French, 1993), the price momentum (Jegadeesh & Titman, 1993), the operating profitability (Fama & French, 2015), and the quality (Asness et al., 2019) amongst others. These anomalies are commonly identified by sorting stocks based on an anomaly variable or using Fama-Macbeth regression. The shortcomings of the former method are significant, despite the simplicity of the approach. Fama & French (2008) noted that "sorts are clumsy for examining the functional form of the relation between average returns and an anomaly variable" (p. 1654).

The objective of this study is to reexamine the anomaly-sorted portfolio returns from the perspective of functional data. A recently developed functional principal component analysis (FPCA) approach is applied in this paper. We decompose the anomaly-sorted returns into orthogonal eigenfunctions based on FPCA. These data-driven functions display the underlying risks and are related to the asset pricing factors.

Although cross-sectional stock returns are observed discretely, the risk is continuous. It is logical to investigate these functional patterns in cross-sectional returns using the framework of the FPCA. Prior researches have provided many factors that affect excess returns (Hou et al., 2020). More recently, Croce et al. (2019) suggest a new influencing factor, i.e., the leading premium. In this article, we focus on the five important factors as in Fama & French (2018), namely, market capitalization (*Size*), book-to-market ratio (*B/M*), profitability (*OP*), investment (*Inv*), and price momentum (*Mom*). We find that the first empirical functional principal component (EFPC) captures the variation of the F-F market factor *Mkt*. Moreover, the monotonicity and curvature found in the second and the third EFPCs, respectively, could be attributed to the underlying cross-sectional risks. An additional finding is that the risks on the cross-section are not always stable. *Inv* and *Size* reversed in the 1970s and the 1980s, respectively.

2. Data and methodology

Monthly anomaly-sorted portfolio returns are obtained from Kenneth R. French's website¹. The sample period is July 1963-July 2020. The starting date corresponds to that in Fama & French (2015). We consider the following five sorting variables: *Size*, *B/M*, *OP*, *Inv*, and *Mom*.

2.1. Smooth the discrete data

Assuming that all the stochastic curves are in the square-integrable Hilbert space $L^2[0, 1]$, the inner product and corresponding norm are defined as $\langle x(u), y(u) \rangle = \int_0^1 x(u)y(u)du$ and $\|x(u)\| = (\int_0^1 x^2(u)du)^{1/2}$, respectively. We aim to smooth the cross-sectional constituents of the discrete observed panel returns $\{r_{u,t}, 1 \leq u \leq N, 1 \leq t \leq T\}$ into the continuous functions $\{r_t(u), 0 \leq u \leq 1, 1 \leq t \leq T\}$.

For each time $t, 1 \leq t \leq T$, the functional data $r_t(u)$ can be represented by the *Karhunen-Loève* Theorem, see Horváth & Kokoszka (2012),

$$r_t(u) = \sum_{m=1}^{\infty} \eta_{t,m} \psi_m(u) \approx \sum_{m=1}^M \eta_{t,m} \psi_m(u), \quad 1 \leq t \leq T \quad (1)$$

¹For more details of the dataset, see Kenneth R. French's website, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

where $\psi_m(u)$ are orthonormal functions in $L^2[0, 1]$.

The most well-known basis functions $\psi_m(u)$ are Fourier bases and B-splines bases, which are mainly used for periodical and non-periodical data, respectively. The number of bases M is essential and determined by the trade-off between smoothness and noises. In this paper, we introduce the ridge regression in Ramsay & Silverman (2005) with a penalty $\text{PEN}_4(r_t) = \int [D^4 r_t(u)]^2 du$, where D^4 is the fourth-order differential operator.

To obtain the fitted curves $\hat{r}_t(u)$ for each time t , we solve the optimization problem,

$$\min \sum_{u=1}^N \left[r_{u,t} - \sum_{m=1}^M \eta_{t,m} \psi_m(u) \right]^2 + \lambda \int [D^4 r_t(u)]^2 du \quad (2)$$

where λ measures the rate of exchange between the capacity of fitting data and the smoothness of $\hat{r}_t(u)$. λ is determined by a *generalized cross-validation* criterion.

2.2. Functional principal component analysis

We assume that $\{r_t(u), 0 \leq u \leq 1\}$ is a stationary sequence in $L^2[0, 1]$ satisfying the moment condition $E \|r_t(u)\|^4 < \infty$. The covariance function of $r_t(u)$ is defined as $c(u, v) = E[r_t(u) - \mu(u)][r_t(v) - \mu(v)]$ where $\mu(u)$ is the functional mean of stationary $\{r_t(u)\}$. $c(u, v)$ is symmetric and non-negative definite. $r_t(u)$ can be decomposed into,

$$r_t(u) = \sum_{m=1}^{\infty} \xi_{t,m} \phi_m(u), \quad 0 \leq u \leq 1, 1 \leq t \leq T \quad (3)$$

where $\xi_{t,m} = \langle r_t(u), \phi_m(u) \rangle = \int_0^1 r_t(u) \phi_m(u) du$.

In practice, the sample estimator of covariance function $c(u, v)$ is given by $\hat{c}_T(u, v) = \frac{1}{T} \sum_{t=1}^T [r_t(u) - \mu_T(u)][r_t(v) - \mu_T(v)]$ where $\mu_T(\cdot) = \frac{1}{T} \sum_{t=1}^T r_t(\cdot)$. The unknown eigenvalues and eigenfunctions of $c(u, v)$ can be estimated by $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq 0$ and $\hat{\phi}_1(u), \hat{\phi}_2(u), \dots$ satisfying,

$$\hat{\lambda}_m \hat{\phi}_m(u) = \int_0^1 \hat{c}_T(u, v) \hat{\phi}_m(v) dv, \quad m = 1, 2, \dots \quad (4)$$

The popular way to decompose the functional observations $r_t(u)$ is to project curves into a finite-dimensional space spanned by $\hat{\phi}_1(u), \hat{\phi}_2(u), \dots, \hat{\phi}_M(u)$ and obtain,

$$\hat{r}_t(u) \approx \sum_{m=1}^M \hat{\xi}_{t,m} \hat{\phi}_m(u), \quad 0 \leq u \leq 1, 1 \leq t \leq T \quad (5)$$

Now the dynamics of functional observations are represented by finite empirical eigenfunctions $\hat{\phi}_m(u)$ and harmonic loadings $\hat{\xi}_{t,m}$, which form the M empirical functional principal components (EFPCs).

3. Empirical results

We examine the value-weighted portfolio returns sorted by the F-F six-factor anomalies $\{r_{u,t}, 1 \leq u \leq N, 1 \leq t \leq T\}$, where u is the sorting group and t is a time index. In the first step, the stocks are divided into ten groups ($N = 10$) and the smoothing process is conducted as follows. At each time t , we set fourteen B-splines bases to functionalize the non-periodical $r_{u,t}$. The smoothing procedure generates $T = 685$ functional curves $\hat{r}_t(u)$ covering 1963-2020.

The next step is to decompose these 685 monthly functional return curves $\{\hat{r}_t(u), 1 \leq t \leq T\}$ with FPCA. $\{\hat{\xi}_{t,m} \hat{\phi}_m(u)\}$ are EFPCs extracted from functions $\hat{r}_t(u)$. Table 1 demonstrates that the

top three components explain more than 90% of the variation of the portfolio returns for all the anomalies. However, the rest of the EFPCs contribute slightly to the return variation, thus we treat them as noises. In the analysis below, we focus on the EFPC_{*i*}, *i* = 1, 2, 3.

Table 1: The percentage of the variation explained by the first eight EFPCs

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Total |
|-------------|--------|--------|-------|-------|-------|-------|-------|-------|--------|
| <i>Size</i> | 94.34% | 3.71% | 0.69% | 0.34% | 0.23% | 0.20% | 0.19% | 0.14% | 99.84% |
| <i>B/M</i> | 87.96% | 4.77% | 2.21% | 1.35% | 1.06% | 0.82% | 0.72% | 0.58% | 99.48% |
| <i>OP</i> | 89.73% | 3.17% | 2.02% | 1.22% | 1.06% | 0.89% | 0.74% | 0.58% | 99.43% |
| <i>Inv</i> | 88.42% | 4.27% | 2.14% | 1.34% | 1.00% | 0.93% | 0.75% | 0.65% | 99.50% |
| <i>Mom</i> | 82.87% | 10.27% | 2.57% | 1.06% | 0.72% | 0.70% | 0.67% | 0.48% | 99.33% |

Note: The variance contribution of empirical eigenfunctions $\hat{\phi}_m$ is captured by $\hat{\lambda}_m$ in Equation 4.

For comparison, the first three empirical eigenfunctions of the return curves for anomaly-sorted portfolios are plotted in Figure 1. The first eigenfunction $\hat{\phi}_1(u)$ slopes slightly for all the anomalies. That is to say, $\hat{\phi}_1(u)$ shows little relation to the sorting criterion. In addition, the monotonicity and curvature in the second and the third EFPCs, respectively, imply a linear and a quadratic relationship with sorting percentage *u*.

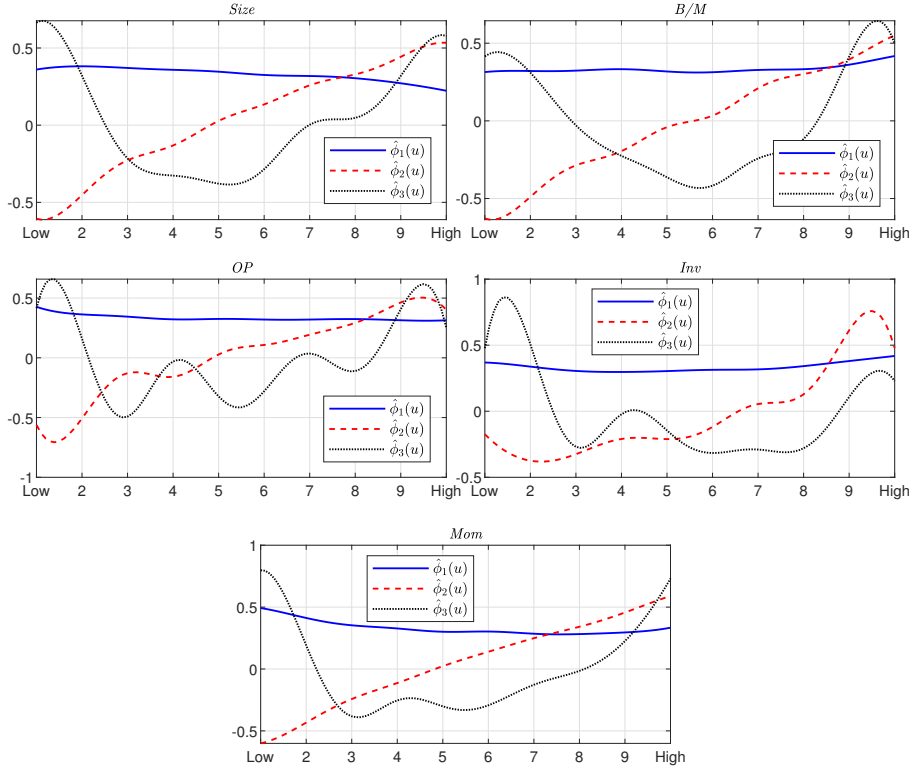


Figure 1: The first three empirical eigenfunctions for all the sorted portfolios

3.1. Market factor along the time-series

The first empirical eigenfunction $\hat{\phi}_1(u)$ exhibits remarkable features on the cross-section. We analyze the expectation of the EFPC₁ with respect to *u* to show how it works.

$$E_u[\hat{\xi}_{t,1}\hat{\phi}_1(u)] = \hat{\xi}_{t,1}E_u[\hat{\phi}_1(u)] \quad (6)$$

The expectation above represents the time-series features of the EFPC₁. The theory is described below. FPCA decomposes the sequence of return curves into orthogonal components. Moreover, the first eigenfunction is weakly affected by the process of cross-sectional sorting. This leads naturally to the time-series analysis. Although the anomalies differ from each other, the expectation of EFPC₁ with respect to u is consistent with the F-F market factor Mkt , see Figure 2.

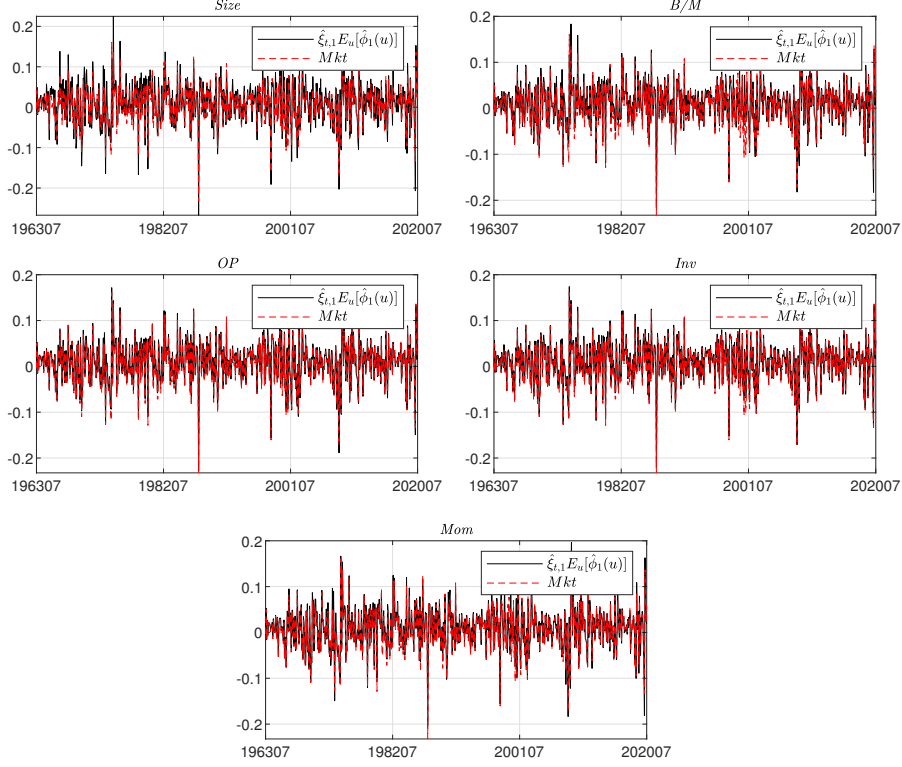


Figure 2: The expectation of the first empirical functional principal component with respect to u

To build upon this, we then regress the expectation of the EFPC₁ on the market factor Mkt ,

$$\hat{\xi}_{t,1}E_u[\hat{\phi}_1(u)] = \beta_0 + \beta_1 Mkt_t + \epsilon_t \quad (7)$$

Table 2: Relationship between the expectation $\hat{\xi}_{t,1}E_u[\hat{\phi}_1(u)]$ and Mkt

| | <i>Size</i> | <i>B/M</i> | <i>OP</i> | <i>Inv</i> | <i>Mom</i> |
|-----------------|---------------------|---------------------|----------------------|----------------------|---------------------|
| $\hat{\beta}_0$ | 0.0044 (6.0489) | 0.0048 (10.7466) | 0.0036 (16.7758) | 0.0046 (24.5872) | 0.0031 (6.8023) |
| $\hat{\beta}_1$ | 1.1135 (68.0977) | 0.9626 (96.5198) | 0.9973 (205.9733) | 0.9755 (234.0488) | 1.0101 (99.2331) |
| adj. R^2 | 0.8714 | 0.9316 | 0.9841 | 0.9877 | 0.9350 |

Note: Bracketed values are t-statistics. All the coefficients are significant at the 1% threshold level.

From the results in Table 2, a clear linear relationship is demonstrated between the expectation $\hat{\xi}_{t,1}E_u[\hat{\phi}_1(u)]$ and the market factor Mkt . The empirical principal component $\hat{\xi}_{t,1}\hat{\phi}_1(u)$ could be regarded as an expansion of Mkt . Regardless of which anomaly is chosen to construct the sorted portfolios, most of the variation in the EFPC₁ along the time-series can be explained by Mkt . Consequently, the EFPC₁ captures the dynamics of the market factor Mkt .

3.2. Pricing factors on the cross-section

Since the EFPC₁ dominates the time-series features, it is necessary to explain the cross-sectional relationships from $\hat{\xi}_{t,2}\hat{\phi}_2(u)$ and $\hat{\xi}_{t,3}\hat{\phi}_3(u)$. We first focus on the EFPC₂. In the previous analysis, empirical eigenfunction $\hat{\phi}_2(u)$ displays the monotonicity for all the anomalies in Figure 1. The positive or negative linearity of $\hat{\xi}_{t,2}\hat{\phi}_2(u)$ with u decisively relies on the sign of $\hat{\xi}_{t,2}$ for each t . In a more feasible way, the *long-run effects* of the EFPC₂ can be represented by the expectation of $\hat{\xi}_{t,2}\hat{\phi}_2(u)$ with respect to t ,

$$E_t[\hat{\xi}_{t,2}\hat{\phi}_2(u)] = E_t[\hat{\xi}_{t,2}]\hat{\phi}_2(u) \quad (8)$$

The long-run effects depend on the eigenfunction $\hat{\phi}_2(u)$ and the expectation of corresponding harmonic loadings $E_t[\hat{\xi}_{t,2}]$. The results for all the anomaly-sorted portfolios are plotted in Figure 3.

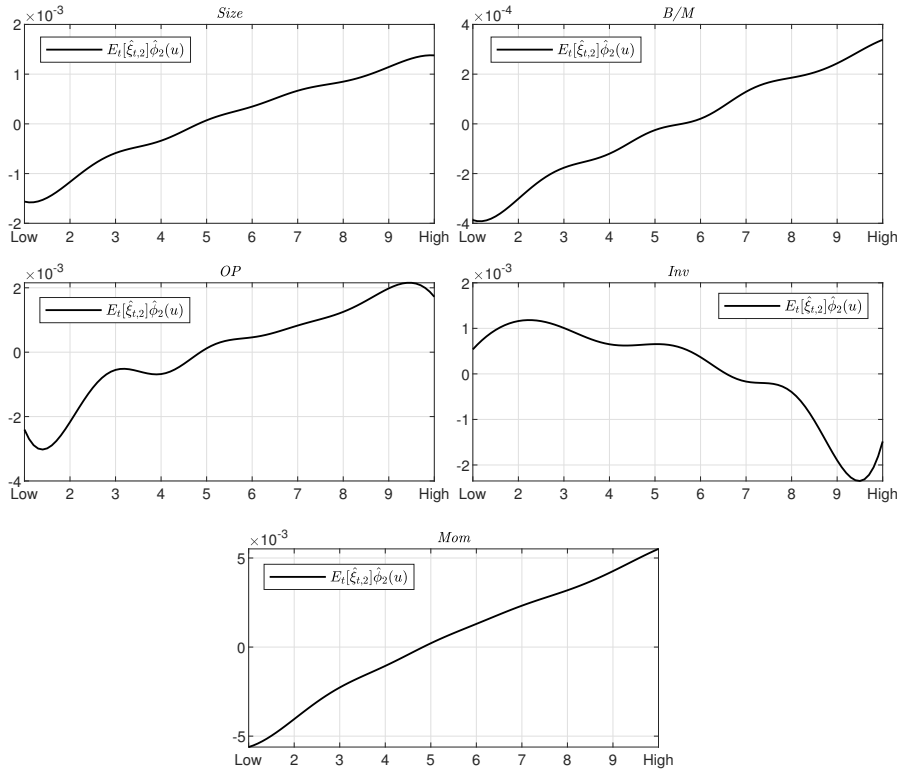


Figure 3: The expectation of the EFPC₂ with respect to t

All empirical results show monotonicity to varying degrees. For instance, the long-run effect of the EFPC₂ for *Inv* is slightly decreasing on the percentage u , which implies the cross-sectional pattern that stocks with a conservative investment strategy obtain higher expected returns than aggressive ones. The remaining four functions are increasing on u . All but *Size* are in accordance with Fama & French (2018).

By equally sampling from the curves with 100 points, we regress $E_t[\hat{\xi}_{t,2}]\hat{\phi}_2(u)$ on u (standardized on $[0, 1]$),

$$E_t[\hat{\xi}_{t,2}]\hat{\phi}_2(u) = \alpha + \beta u + \epsilon_u \quad (9)$$

The results in Table 3 confirm the linear relationship between $E_t[\hat{\xi}_{t,2}]\hat{\phi}_2(u)$ and u . $\hat{\phi}_2(u)$ is the only eigenfunction containing a linear relationship and the EFPC₂ accounts for the highest

Table 3: Relationship between the expectation $E_t[\hat{\xi}_{t,2}]\hat{\phi}_2(u)$ and u

| | <i>Size</i> | <i>B/M</i> | <i>OP</i> | <i>Inv</i> | <i>Mom</i> |
|----------------|----------------------|----------------------|-----------------------|-----------------------|-----------------------|
| $\hat{\alpha}$ | -0.0014 (-56.302) | -0.0004 (-87.989) | -0.0023 (-31.6143) | 0.0017 (18.2335) | -0.0049 (-74.1476) |
| $\hat{\beta}$ | 0.0029 (69.7142) | 0.0007 (99.3374) | 0.0047 (37.9013) | -0.0032 (-20.1909) | 0.0106 (93.2222) |
| adj. R^2 | 0.9800 | 0.9901 | 0.9355 | 0.8042 | 0.9887 |

Note: Bracketed values are t-statistics. All the coefficients are significant at the 1% threshold level.

proportion of the variation of returns, except for the market EFPC₁. Consequently, if there exists any cross-sectional linear relationship, $\hat{\xi}_{t,2}\hat{\phi}_2(u)$ captures it.

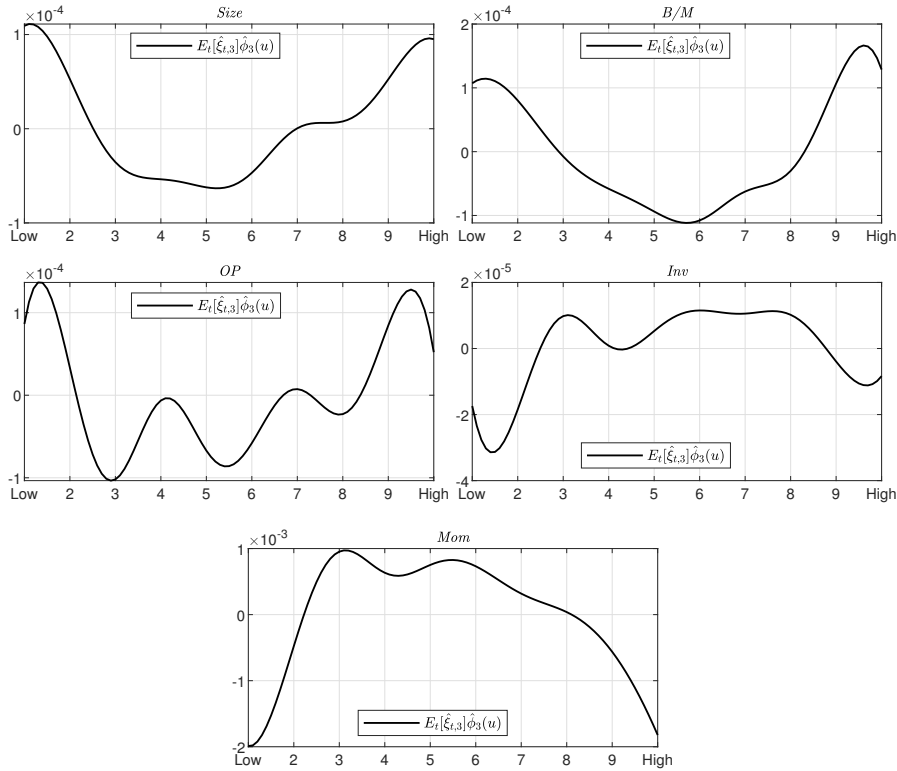


Figure 4: The expectation of the EFPC₃ with respect to t

The long-run effects of the EFPC₃ show quadratic patterns, see Figure 4. To test this, we regress $E_t[\hat{\xi}_{t,3}]\hat{\phi}_3(u)$ on u and u^2 ,

$$E_t[\hat{\xi}_{t,3}]\hat{\phi}_3(u) = \alpha_0 + \alpha_1 u + \alpha_2 u^2 + \epsilon_u \quad (10)$$

The curvature of the expectation $E_t[\hat{\xi}_{t,3}]\hat{\phi}_3(u)$ implies the different attitudes of the market towards the extremes and middles. Barber & Odean (2008) propose the *attention-grabbing trading hypothesis*. The consideration set of investors only contains the attention-grabbing stocks, i.e., stocks meeting extreme criteria. It is likely that the curvature of the EFPC₃ reflects the limited concentration of investors in markets.

Table 4: Relationship between the expectation $E_t[\hat{\xi}_{t,3}]\hat{\phi}_3(u)$ and u, u^2

| | <i>Size</i> | <i>B/M</i> | <i>OP</i> | <i>Inv</i> | <i>Mom</i> |
|------------------|-----------------------|-----------------------|-----------------------|------------------------|-----------------------|
| $\hat{\alpha}_0$ | 0.0001 (20.2455) | 0.0002 (30.5843) | 0.0001 (7.8666) | -0.00003 (-17.2844) | -0.0015 (-15.3746) |
| $\hat{\alpha}_1$ | -0.0007 (-27.6413) | -0.0011 (-42.1656) | -0.0007 (-12.1131) | 0.0001 (18.9258) | 0.0099 (22.1703) |
| $\hat{\alpha}_2$ | 0.0007 (29.5452) | 0.0011 (44.3128) | 0.0007 (13.4072) | -0.0001 (-17.4401) | -0.0102 (-23.6036) |
| adj. R^2 | 0.8994 | 0.9521 | 0.6567 | 0.7881 | 0.8503 |

Note: Bracketed values are t-statistics. All the coefficients are significant with level of 0.01.

3.3. Two special cases: *Size* and *Inv*

Recall that the cross-sectional long-run effects $E_t[\hat{\xi}_{t,2}]\hat{\phi}_2(u)$ are determined by the sign of $E_t[\hat{\xi}_{t,2}]$. The expectation of $\hat{\xi}_{t,2}$ is estimated by the sample mean which is related to the partial sum of $\hat{\xi}_{t,2}$. Figure 5 plots the partial sum for all the anomalies. For the majority of the sample period, $E_t[\hat{\xi}_{t,2}]$ of anomalies *B/M*, *OP*, and *Mom* remains positive and thus the long-run effects of these anomalies are stable over time.

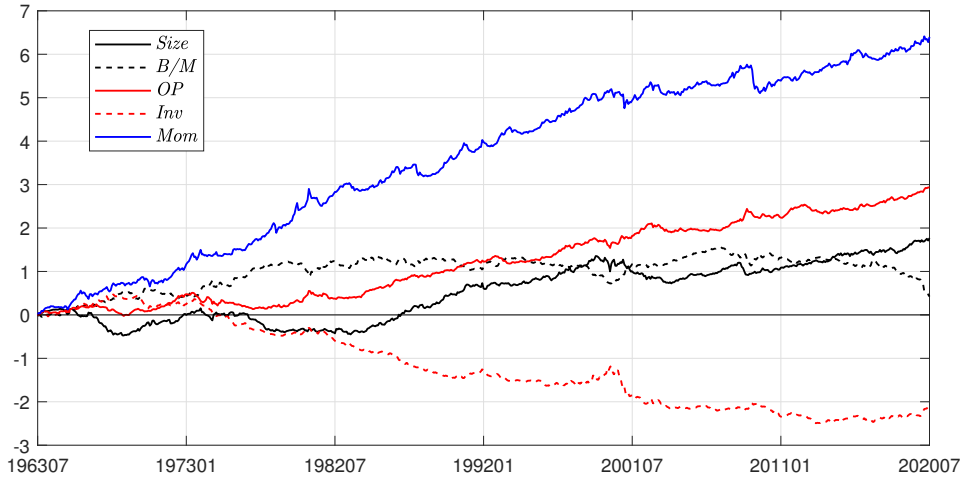


Figure 5: Partial sum of $\hat{\xi}_{t,2}$

We here illustrate two special cases, *Size* and *Inv*. In subsection 3.2, the long-run effect of EFPC₂ for *Size* are upward-sloping. That is to say stocks with a large market cap have higher expected returns. It is clear that this conclusion is in contrast with the pricing factor *Size* in Fama & French (1993). The reasons are as follows. Partial sum of $\hat{\xi}_{t,2}$ for *Size* is negative for the most period of time before the 1980s and reverses after that, see Figure 5. Combined with the upsloping eigenfunction $\hat{\phi}_2(u)$ in Figure 1, the long-run effect $E_t[\hat{\xi}_{t,2}]\hat{\phi}_2(u)$ of *Size* slopes downwards alongside the percentage of portfolios u in the first period and upwards in the second.

As proposed by Fama & French (2015), the *Inv* factor shows the relative profitability of firms with a conservative investment strategy. Using the framework of this time-series analysis, the neglected reversal in the 1970s is highly apparent.

4. Conclusion

This paper introduces the functional principal component analysis approach for decomposing the panel returns for portfolios sorted by *Size*, *B/M*, *OP*, *Inv*, and *Mom*. For all anomalies, the first EFPC captures more than 80% of the total variations and captures the market factor *Mkt*. The second and the third EFPCs display monotonicity and curvature on the cross-section, respectively. We consider these to be the cross-sectional underlying risks. Moreover, two remarkable reversals in the monotonicity of *Inv* and *Size* occurred in the 1970s and the 1980s.

Previous studies have shown that, regardless of which anomaly is chosen to construct the portfolios, the market EFPC dominates the variation of returns. This is the reason why so many asset pricing models contain the market factor *Mkt* due to its important role (Sharpe, 1964; Fama & French, 1993, 2015, 2018; Carhart, 1997). Furthermore, FPCA provides us with insight into the time-series structure of cross-sectional anomalies.

References

- Asness, C. S., Frazzini, A., & Pedersen, L. H. (2019). Quality minus junk. *Review of Accounting Studies*, 24, 34–112.
- Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies*, 21, 785–818.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, 52, 57–82.
- Croce, M. M., Marchuk, T., & Schlag, C. (2019). *The leading premium*. Technical Report National Bureau of Economic Research.
- Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, .
- Fama, E. F., & French, K. R. (2008). Dissecting anomalies. *The Journal of Finance*, 63, 1653–1678.
- Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, 116, 1–22.
- Fama, E. F., & French, K. R. (2018). Choosing factors. *Journal of financial economics*, 128, 234–252.
- Horváth, L., & Kokoszka, P. (2012). *Inference for functional data with applications* volume 200. Springer Science & Business Media.
- Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. *The Review of Financial Studies*, 33, 2019–2133.
- Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48, 65–91.
- Ramsay, J. O., & Silverman, B. W. (2005). *Functional Data Analysis*. Springer.
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19, 425–442.