How Robust are Empirical Factor Models to the Choice of Breakpoints?

Fabian Hollstein, Marcel Prokopczuk, and Victoria Voigts

Leibniz University Hannover

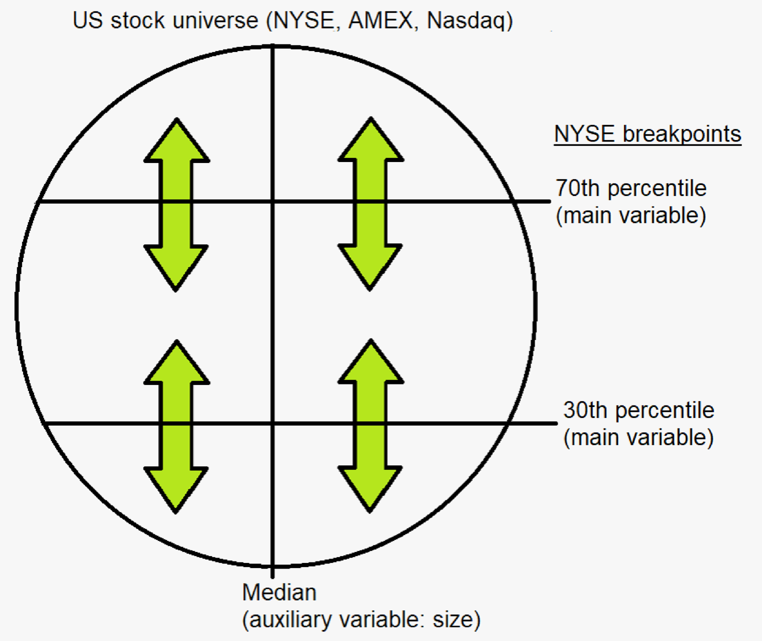
FULL PAPER

Abstract

We comprehensively investigate the robustness of well-known factor models to altered factor-formation breakpoints. Deviating from the standard $30th$ and $70th$ percentile selection, we use an extensive set of anomaly test portfolios to uncover two main findings: On the one hand, there is a trade-off between specification versus diversification. More centered breakpoints tend to result in less (idiosyncratic) risk. More extreme sorts create stronger exposures to the underlying anomalies and, thus, higher average returns. On the other hand, the models are robust to different degrees. The Hou–Xue–Zhang [6] model is much more sensitive to changes in breakpoints than the Fama–French models.

Introduction & Motivation

- Fama and French (1993) [2] pioneered the use of firm characteristics to create empirical factor models.
- They define a benchmark procedure that breaks the main variable at the $30th$ and $70th$ percentile, whereas the auxiliary variable is split at the median.



- However, this specific choice is neither economically motivated nor the result of an optimization problem:

The splits are arbitrary, however, and we have not searched over alternatives. The hope is that the tests here and in Fama and French (1992)[1] are not sensitive to these choices. We see no reason to argue that they are. — Fama and French (1993, p.9)

Literature

The growing literature that examines the robustness of documented asset pricing phenomena shows that certain methodological choices can already have a considerable impact. Thereby, previous studies focus on the effect of...

- the rebalancing frequency (Liew and Vassalou, 2000) [10]
- the mean–variance–optimal combinations of the factor base portfolios (Grinblatt and Saxena, 2018) [4]
- using whole-sample vs. NYSE–breakpoints with equal- or value-weighting (Hou, Xue, and Zhang, 2020) [7]
- dependent vs. independent sorting (Lambert, Fays, and Hübner, 2020) [8]

Brief attempts addressing the effect of shifting breakpoints in the factor formation process do not present a homogeneous picture:

- Fama and French (2015) [3]: find that the standard 2×3 double sort, the double 2×2, and the quadruple $2\times2\times2\times2$ sorts with median breakpoints *"produce much the same results"* (p. 10)
- Li, Novy-Marx, and Velikov (2019) [9]: document that the Pastor and Stambaugh (2003) [12] liquidity factor is very sensitive to shifts from the authors' proposed scheme to the Fama–French benchmark sorting methodology, leading to a substantially diminished performance.

Research Questions

1. Does deviating from the standard breakpoint choice (30–70) have an influence on the factor models' performances?
   → **Are the pricing results robust to this choice?**
2. Do different models vary in their sensitivity towards changes in construction or exhibit different levels of robustness?
   → **Do some models perform even better for alternative choices?**

Methodology

We complement existing studies on the robustness of documented asset pricing phenomena by systematically analyzing the impact of the quantile-breakpoint selection on the factor and model performance.

Table 1. **Overview of the Standard Factor Construction Procedures**
This table provides an overview of the standard construction procedure of each factor considered in this paper (i.e., SMB, HML, RMW, CMA, UMD, LIQ, SMB_q, ROE_q, and IA_q). Beside the original author(s) and the dimensions of the sort, we present the sorting variable(s) and breakpoint choices for both the main and auxiliary variable(s).

Factor	Author(s)	Sorting Variable		Breakpoints		Sort
		Main	Auxiliary	Main	Auxiliary	
SMB	Fama and French (1993)	size	value/operating profitability/investment	50–50	30–70	double
HML	Fama and French (1993)	value	size	30–70	50–50	double
RMW	Fama and French (2015)	operating profitability	size	30–70	50–50	double
CMA	Fama and French (2015)	investment	size	30–70	50–50	double
UMD	Carhart (1997)/ Fama and French (2010a)	prior 12-months returns	none/ size	30–70	none/ 50–50	single/ double
SMB_q	Hou, Xue, and Zhang (2015)	size	return on equity/ investment	50–50	30–70/ 30–70	triple
ROE_q	Hou, Xue, and Zhang (2015)	return on equity	size/ investment	30–70	50–50/ 30–70	triple
IA_q	Hou, Xue, and Zhang (2015)	investment	size/ return on equity	30–70	50–50/ 30–70	triple
LIQ	Pástor and Stambaugh (2003)	liquidity	none/ size	10–90	none/ 50–50	single/ double

Data and Variables

Factor Models

$$\text{FF3:}\quad R_{it}=\alpha_i+\beta_i MktRF_t+s_i SMB_t+h_i HML_t+\varepsilon_{it}$$
$$\text{C4:}\quad R_{it}=\alpha_i+\beta_i MktRF_t+s_i SMB_t+h_i HML_t+u_i UMD_t+\varepsilon_{it}$$
$$\text{PS4:}\quad R_{it}=\alpha_i+\beta_i MktRF_t+s_i SMB_t+h_i HML_t+l_i LIQ_t+\varepsilon_{it}$$
$$\text{CPS5:}\quad R_{it}=\alpha_i+\beta_i MktRF_t+s_i SMB_t+h_i HML_t+l_i LIQ_t+u_i UMD_t+\varepsilon_{it}$$
$$\text{FF5:}\quad R_{it}=\alpha_i+\beta_i MktRF_t+s_i SMB_t+h_i HML_t+r_i RMW_t+c_i CMA_t+\varepsilon_{it}$$
$$\text{FF6:}\quad R_{it}=\alpha_i+\beta_i MktRF_t+s_i SMB_t+h_i HML_t+r_i RMW_t+c_i CMA_t+u_i UMD_t+\varepsilon_{it}$$
$$\text{HXZ4:}\quad R_{it}=\alpha_i+\beta_i MktRF_t+s_i SMB_t+r_i ROE_t+c_i IA_t+\varepsilon_{it}$$

Data Sources

The selected sample period is January 1973 to September 2018 to ensure continuous data availability of all constructed factors. → Increasing the number of portfolios, which the stocks are (unevenly) assigned to, increases the likelihood of empty portfolios. This is especially acute during times when the total number of stocks is small or for breakpoint choices that are shifted towards the edges of the distribution.

- CRSP (daily and monthly)
- Compustat (quarterly and annual)
- Hou-Xue-Zhang q-factors Data Library
- Ken French Data Library

LHS Portfolios

We use 185 anomaly variables among 6 categories: momentum (40), value-versus-growth (32), investment (29), profitability (44), intangibles (30), and trading frictions (10) taken from Hou-Xue-Zhang q-factors Data Library

- **Main analysis:** single sorted decile portfolios are formed into long–short portfolios
  → one portfolio for each anomaly
- **Robustness analysis:** independent double (3×5) sorts, where the anomaly interacts with size
  → 15 portfolios for each anomaly

Asset Pricing Metrics

- $A|\alpha|$: Mean absolute α
- $\#^{H-L}$: Number of significant long–short portfolios (for long–short LHS portfolios)
- $\#GRS$: Number of rejected GRS-tests for each set of LHS portfolios (for a set of LHS portfolios)
- $\frac{A\alpha_i^2}{Ar_i^2}$: Fraction of total dispersion that is due to α-dispersion (for a set of LHS portfolios)
- $\frac{As^2(\alpha_i)}{A\alpha_i^2}$: Fraction of α-dispersion that is due to noise (for a set of LHS portfolios)
- $A\bar{R}^2$: Mean adjusted R^2

Main Analysis

At the factor level, we find an overall pattern in favour of breakpoints shifted towards the outside margin on the factor level in several analyses. However, we could not confirm this on the model (or tangency portfolio) level regarding maximum model Sharpe ratios. → *see full paper*

For insights into the performances of differently specified models, we further examine their performance in standard asset pricing tests:

Table 2. **Summary Tests of Asset Pricing Models for 185 Sets of Anomaly Long–Short Portfolios at 0.27% Significance Level**
We compare the asset pricing performance for all breakpoint specifications of each factor model considered in this study (i.e., FF3, C4, PS4, CPS5, FF5, FF6, and HXZ4). The anomalies considered as LHS long–short regression portfolios are the momentum (*Mom*), value-versus-growth (*V–G*), investment (*Inv*), profitability (*Prof*), intangibles (*Intan*), frictions (*Fric*) categories, as well as all anomalies from all categories combined (*All*). The anomaly returns are all downloaded from the Hou–Xue–Zhang q-factors Data Library website. The numbers in parentheses are the numbers of long–short anomaly portfolios (one for each anomaly) in each category. $A|\alpha^{H-L}|$ is the average magnitude of the high-minus-low alphas. Following the recommendations of Harvey et al. (2016) [5], we only count the number of high-minus-low alphas ($\#^{H-L}$) with t-statistic $\geq|3|$ corresponding to a significance **at the 0.27% level**. A darker shade of green indicates a relatively better performing specification w.r.t. the asset pricing metric for each model within each category. Standard errors are adjusted for heteroskedasticity and autocorrelation following Newey and West (1994) [11] by using an automatic lag selection. All returns are annualized and in percentage points. The sample period ranges from January 1973 to September 2018 (549 months).

Model	All (185)		Mom (40)		V-G (32)		Inv (29)		Prof (44)		Intan (30)		Fric (10)	
	$A\lvert\alpha^{H-L}\rvert$	$\#^{H-L}$	$A\lvert\alpha^{H-L}\rvert$	$\#^{H-L}$	$A\lvert\alpha^{H-L}\rvert$	$\#^{H-L}$	$A\lvert\alpha^{H-L}\rvert$	$\#^{H-L}$	$A\lvert\alpha^{H-L}\rvert$	$\#^{H-L}$	$A\lvert\alpha^{H-L}\rvert$	$\#^{H-L}$	$A\lvert\alpha^{H-L}\rvert$	$\#^{H-L}$
$FF3_{50-50}$	5.86	99	7.74	34	1.77	2	3.53	8	8.87	38	5.97	13	4.67	4
$FF3_{40-60}$	5.88	95	7.76	34	1.78	2	3.65	8	8.86	36	5.90	11	4.86	4
$FF3_{30-70}$	5.87	97	7.74	34	1.80	2	3.62	8	8.82	38	5.94	11	4.85	4
$FF3_{20-80}$	5.81	96	7.65	33	1.81	2	3.49	8	8.67	36	6.03	13	4.62	4
$FF3_{10-90}$	5.62	90	7.29	32	1.76	2	3.64	6	7.94	31	6.44	15	4.42	4
$C4_{50-50}$	4.30	60	3.06	6	2.50	5	2.92	6	6.78	28	5.81	14	3.55	1
$C4_{40-60}$	4.31	56	3.00	5	2.65	5	3.01	6	6.72	27	5.76	12	3.69	1
$C4_{30-70}$	4.21	53	2.80	5	2.65	5	2.94	6	6.51	26	5.83	10	3.56	1
$C4_{20-80}$	4.04	53	2.49	3	2.68	5	2.79	6	6.15	27	5.93	11	3.21	1
$C4_{10-90}$	4.00	49	2.42	2	3.20	7	2.95	5	5.43	22	6.22	12	3.03	1
$PS4_{50-50}$	6.05	101	7.94	35	1.86	2	3.64	7	9.02	39	6.37	14	4.83	4
$PS4_{40-60}$	6.09	102	8.06	37	1.78	2	3.75	8	9.08	37	6.28	14	4.99	4
$PS4_{30-70}$	6.10	100	8.09	35	1.82	2	3.71	7	9.10	38	6.29	14	4.96	4
$PS4_{20-80}$	6.00	94	7.92	33	1.82	2	3.57	6	8.93	36	6.33	13	4.78	4
$PS4_{10-90}$	5.58	87	7.24	31	1.57	1	3.61	5	7.90	33	6.56	14	4.30	3
$CPS5_{50-50}$	4.34	61	3.02	7	2.26	3	3.01	7	6.86	29	6.11	13	3.68	2
$CPS5_{40-60}$	4.34	59	2.94	6	2.37	4	3.09	7	6.82	28	6.08	12	3.75	2
$CPS5_{30-70}$	4.26	57	2.78	6	2.40	3	3.00	6	6.68	28	6.12	13	3.59	1
$CPS5_{20-80}$	4.11	53	2.51	5	2.50	5	2.85	4	6.32	27	6.19	12	3.32	0
$CPS5_{10-90}$	3.98	45	2.48	1	2.93	6	2.93	3	5.42	21	6.31	13	3.00	1
$FF5_{50-50}$	4.64	64	7.17	24	1.47	1	2.54	5	5.92	24	5.49	9	2.64	1
$FF5_{40-60}$	4.50	64	6.87	23	1.48	1	2.52	6	5.61	24	5.45	9	2.67	1
$FF5_{30-70}$	4.45	58	6.85	20	1.39	1	2.56	4	5.43	24	5.57	8	2.35	1
$FF5_{20-80}$	4.50	61	6.95	21	1.35	1	2.54	4	5.53	25	5.67	8	2.33	2
$FF5_{10-90}$	4.57	68	6.71	22	1.35	1	2.47	5	5.96	29	5.70	9	2.89	2
$FF6_{50-50}$	3.30	40	3.08	6	1.88	1	2.28	4	4.22	17	5.23	11	1.88	1
$FF6_{40-60}$	3.26	40	2.98	6	2.00	2	2.28	4	4.02	16	5.25	11	1.93	1
$FF6_{30-70}$	3.20	36	2.88	6	1.92	2	2.30	3	3.84	13	5.37	11	1.79	1
$FF6_{20-80}$	3.15	35	2.64	5	1.97	2	2.25	3	3.78	13	5.42	11	1.93	1
$FF6_{10-90}$	3.27	40	2.50	4	2.43	4	2.22	3	3.90	16	5.62	11	2.20	2
$HXZ4_{50-50}$	3.45	29	3.51	5	2.84	0	2.59	3	3.44	10	5.41	9	1.81	2
$HXZ4_{40-60}$	3.25	28	3.12	4	2.34	0	2.42	3	3.60	11	5.23	8	1.68	2
$HXZ4_{30-70}$	3.09	25	2.87	3	2.05	0	2.43	3	3.32	8	5.22	9	1.82	2
$HXZ4_{20-80}$	3.17	26	2.70	3	2.56	0	2.44	3	3.26	8	5.43	10	1.91	2
$HXZ4_{10-90}$	4.18	37	3.66	5	4.86	2	3.55	8	3.41	9	6.22	10	3.15	3

Results

- More successful models in explaining average returns (i.e., FF5, FF6, and HXZ4) work best in the moderate breakpoint area centered around 30–70. These models perform worst for 10–90 sorts, followed by median splits.
- Models including the UMD factor (i.e., C4 and CPS5) perform best for 10–90 breakpoint choices and gradually deteriorate for other specifications.
- The LIQ factor only adds little explanatory power to the models.
- At the 0.27% significance level, we receive clear performance patterns for different model specifications (contradicting patterns in performance metrics for more simplistic models (i.e., FF3 and PS4) at lower significance threshold such as 5% or 10% → *see full paper*).
- The FF5 and FF6 model appear to be more robust than the HXZ4 model: breakpoints can moderately be shifted in either direction (20–80 or 40–60) without suffering noticeable detriment in model performance.
- The HXZ4 model, however, reacts more heavily to deviations in breakpoints.

Robustness Analyses

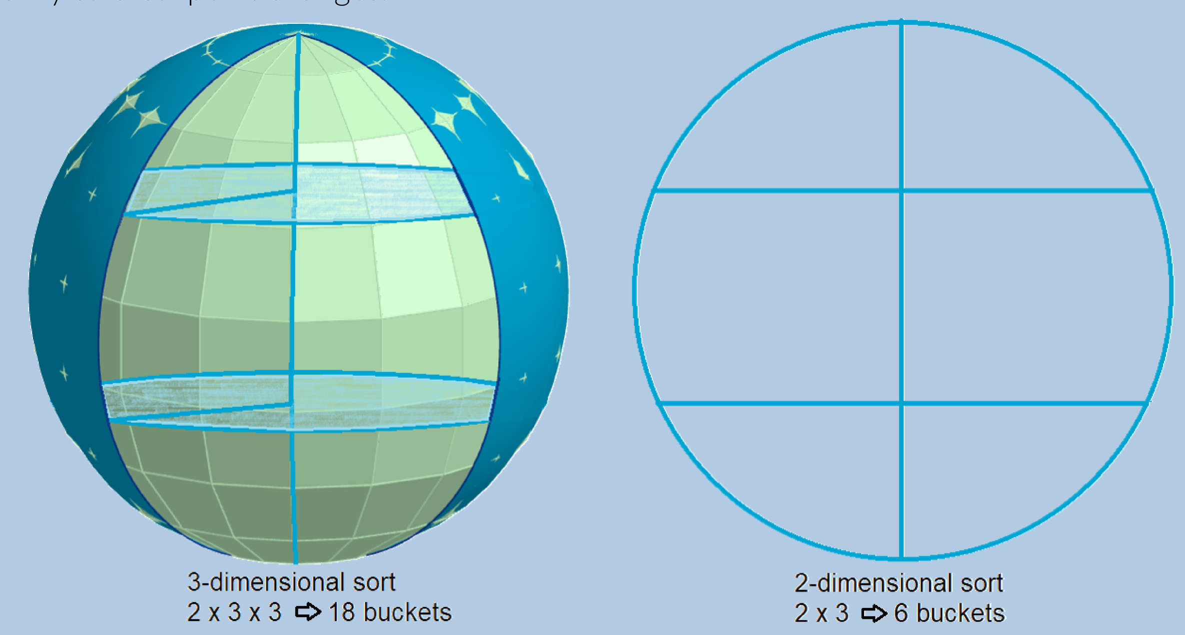
We repeat our main analysis further incorporating alternative...

- **significance levels**: We consider a set of different significance hurdles in our analyses. On the one hand, we consider the recommendations of Harvey, Liu, and Zhu (2016) [5] by implementing a much higher than standard 5% hurdle, where the $|t\text{-statistic}|\geq3$, corresponding to a significance at the 0.27% level. On the other hand, we check our results at the 10% significance level. We overall find the same patterns in the regression output, however, the higher the significance hurdle, the clearer the performance patterns for different model specifications. This holds especially true for more simplistic models (e.g., FF3, PS4).
- **data sorts**: For our main analysis, we intentionally use long–short LHS portfolios: Since they exhibit the largest possible spread in the anomaly, they are the most difficult to price (and thus the most challenging for the factor models). Being easier to price, double sorted sets of LHS portfolios generate results that are noticeably more homogeneous for all models across all anomaly categories. Our findings are supported by running the analysis with different sets of LHS portfolios.
- **asset pricing metrics**: For our robustness analysis, we incorporate $\#GRS$, $\frac{A\alpha_i^2}{A\bar{r}_i^2}$, and $\frac{As^2(\alpha_i)}{A\alpha_i^2}$. Overall, all asset pricing metrics support the same performance pattern.
- **sets of LHS data**: We further verify our results by repeating the analyses using 15 long–short and 14 double 5×5 sorts provided by Ken French's Data Library.

Findings

We contribute to the empirical asset pricing literature by conducting a comprehensive robustness analysis on the impact of breakpoint changes in factor construction. We obtain two major results.

1. The choice of more centered breakpoints tends to result in less (idiosyncratic) risk at the cost of lower expected returns. Consequently, we observe a **trade-off between specification versus diversification**. More extreme sorts exhibit larger exposures to the anomalies at the price of a lack of diversification, whereas factor models constructed using an intermediate breakpoint choice such as 30–70 tend to profit from less idiosyncratic noise while the exposures to the anomalies are smaller.
2. The models exhibit **different levels of robustness:** Due to its construction where all stocks are potentially unevenly split into 18 formation portfolios, the HXZ4 model reacts more heavily to breakpoint changes.



Conversely, the majority of Fama–French based models turn out to be fairly robust for slight shifts in breakpoints (typically around the range median or 40–60 to 20–80).

Guidelines

Taking all our observations into consideration, we generally advocate for choosing more centered breakpoints (i.e., $30th$ and $70th$ percentile and $20th$ and $80th$ percentile) in order to achieve best model performances. However, we find deviations from that general rule and identify three potential guidelines:

1. **Dimension of the sort**: a higher dimension sort allocates the total number of stocks into a greater number of portfolios. Thus, it is more intensely prone to idiosyncratic noise when breakpoints are shifted towards the edges of the distribution. For these types of models more centered breakpoints are more appropriate.
2. **Data quality**: If the cross-section fails to be sufficiently large at each point in time, same applies as above and more centered breakpoints are advisable.
3. **Complexity of the model**: more complex models tend to capture anomaly return patterns more successfully than more parsimonious models. Here, a higher anomaly exposure in factor construction does not add as much explanatory power to the model and thus, does not outweight the performance diminishing effects of the added noise. Hence, more complex the models tend to work best for a more centered breakpoint choice.

References

[1] Eugene F. Fama and Kenneth R. French. The Economic Fundamentals of Size and Book-to-Market Equity. Working Paper (Graduate School of Business. University of Chicago), 1992b.

[2] Eugene F. Fama and Kenneth R. French. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.

[3] Eugene F. Fama and Kenneth R. French. A Five-Factor Asset Pricing Model. *Journal of Financial Economics*, 116(1):1–22, 2015.

[4] Mark Grinblatt and Konark Saxena. When Factors Do Not Span Their Basis Portfolios. *Journal of Financial and Quantitative Analysis*, 53(6):2335–2354, 2018.

[5] Campbell R. Harvey, Yan Liu, and Heqing Zhu. ... and the cross-section of expected returns. *Review of Financial Studies*, 29(1):5–68, 2016.

[6] Kewei Hou, Chen Xue, and Lu Zhang. Digesting Anomalies: An Investment Approach. *Review of Financial Studies*, 28(3):650–705, 2015.

[7] Kewei Hou, Chen Xue, and Lu Zhang. Replicating Anomalies. *Review of Financial Studies*, 33(5):2019–2133, 2020.

[8] Marie Lambert, Boris Fays, and Georges Hübner. Factoring Characteristics into Returns: A Clinical Study on the SMB and HML Portfolio Construction Methods. *Journal of Banking and Finance*, 114(105811):1–20, 2020.

[9] Hongtao Li, Robert Novy-Marx, and Mihail Velikov. Liquidity Risk and Asset Pricing. *Critical Finance Review*, 8(1-2):223–255, 2019.

[10] Jimmy Liew and Maria Vassalou. Can Book-to-Market, Size and Momentum be Risk Factors that Predict Economic Growth? *Journal of Financial Economics*, 57(2):221–245, 2000.

[11] Whitney K. Newey and Kenneth D. West. Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies*, 61(4):631–653, 1994.

[12] Lubos Pástor and Robert F. Stambaugh. Liquidity Risk and Expected Stock Returns. *Journal of Political Economy*, 111(3):642–685, 2003.

http://global-q.org/testingportfolios.html
https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

voigts@fcm.uni-hannover.de