Local, Regional, and Global Asset Pricing?

Joshua Traut*†

July 30, 2024

Abstract

Uncovering the underlying structure of global factor returns and assessing whether assets are priced on a local, regional, or global level are important tasks to understand the dynamics of asset pricing. I am the first to assess the regional extent of factor dynamics or the optimal level of aggregation by comprehensively analyzing factor dependencies in 35 countries. Following a data-driven approach I identify three-factor regions along geographical and economic lines. With regards to asset pricing, I grant novel insights that the performance of local asset pricing models is largely driven by the local market factor and that optimal models contain local, regional, and global factors, challenging current findings that local models perform best. The findings offer guidance for international asset pricing tests, deepen understanding of factor return dynamics, and provide evidence of the efficacy of global pricing models.

Keywords: Factor Models, Financial Markets, International Asset Pricing, Market Integration, Cluster Analysis.

JEL Classification: C52, G11, G12, G15, G17.

*University of St. Gallen, Swiss Institute of Banking and Finance, Unterer Graben 21, 9000 St. Gallen, Switzerland | joshua.traut@unisg.ch | https://www.joshuatraut.com

†I am grateful for insightful comments from Merlin Bartel, Niclas Käfer, Manuel Mezger, Paul Söderlind, Kevin Schneider, Paul Schneider, and Tobias Wiest as well as from participants of the Finance Research Seminar at the University of St. Gallen on an earlier version of this paper. All errors are my own.

1 Introduction

In the dynamic landscape of factor asset pricing, the ongoing discourse surrounding the efficacy of local, regional, or global asset pricing factors has been a key point of the discussion. A recent contribution by Hollstein (2022) aimed to settle this debate, offering a comprehensive analysis across 48 equity markets. Analyzing multiple factor models Hollstein (2022) asserts the superiority of local asset pricing models in effectively pricing the cross-section of stocks. While his findings present compelling evidence in favor of local models, some aspects about the evaluation of local vs. regional vs. global asset pricing models remain open, prompting a revisit of the subject.

One notable gap in the existing literature pertains to the level of regionality with which regional asset pricing models are built. Previous studies on regional asset pricing models predominantly employ a broad categorization of regions based on continental affiliations or developmental statuses (Brooks & Del Negro, 2005; Fama & French, 2012, 2017). Recent studies (e.g., Hollstein, 2022; Karolyi & Wu, 2018) adopt the regions defined by Fama and French (2012) who argue that in their regions markets can be assumed to be reasonably integrated. However, they do not explain how they develop those specific regions. Yet, the optimal delineation of regions remains uncharted territory. Should regions encompass only neighboring countries or span entire continents? Can we even find regional integration for asset pricing factors that warrant a rationale to test regional models? Moreover, the studies above define the same regions for all factors they test. But what if the level of regional integration differs among factors? Addressing these gaps, this paper is the first to establish data-driven regions for factor asset pricing analyses.

Furthermore, the extant literature predominantly evaluates asset pricing models based on traditional metrics like R^2 and regression alphas (Fama & French, 2012, 2017; Griffin, 2002; Hollstein, 2022; Hou et al., 2011). However, recent advancements advocating the assessment of asset pricing models through Sharpe ratios (Barillas et al., 2020; Barillas & Shanken, 2017) have yet to be incorporated into this discourse. It is recommended to use Sharpe ratios over traditional alpha intercepts because this measure does not depend on the test assets (Fama & French, 2018). Because of this, more recent papers have started evaluating factor models solely using Sharpe ratios as the evaluation metric (e.g., Detzel et al., 2023). Additionally, the prevailing approach often confines the analysis to examining local, regional, and global models in isolation, neglecting the possibility of hybrid models that encompass both local and global components. Addressing these deficiencies, this study not only evaluates asset pricing models through established metrics but also integrates Sharpe ratios into the assessment framework. Moreover, it explores the possibility of integrated models, composed of both local and global factors and thereby provides a more nuanced understanding of asset pricing dynamics.

To address these research gaps, I adopt a data-driven approach utilizing global factor data that encompasses monthly factor returns across 35 countries from January 1992 to December 2023. For the analysis, I employ the Fama and French (2018) six-factor model that has been shown to be superior over other factor models (Barillas et al., 2020; Fama & French, 2018). However, I also show that the results are robust for other models such as the Fama and French (2015) five-factor model or Hou et al. (2015) q-factor model.

To identify candidate regions for regional asset pricing analyses, I employ principal component analysis (PCA) on the correlation matrices of international factor returns. This investigation reveals distinct groupings of countries by geographic regions and development status. Although the groupings are consistent across factors, their strengths vary significantly between them. An integrated analysis of all factor returns yields that factor returns can be clustered into three regions: an ”established” region encompassing European countries, the U.S., and Canada; an ”emerged” region comprising rising Asian countries like South Korea and Thailand; and a ”developing” region including developing countries from Europe, Asia, and South America. Notably, I do not find support for treating Japan as an outlier, a practice commonly employed in other studies such as Fama and French (2012) and Hollstein (2022).

For the asset pricing analysis I build three different factor models for each country, c,—a local model consisting of factors in c, a regional model that holds factors built from all other countries in c's region, and a rest-of-world model that is constructed from all other countries outside of c's region. Beginning with an evaluation of whether regional factor returns encapsulate those of individual countries, I regress local factor returns on their corresponding regional returns. Remarkably, local factor returns are comprehensively represented by their regional counterpart. Subsequently, I assess the efficacy of local, regional, and rest-of-world factor models in pricing GICS 10 industry portfolios. While local models perform well in terms of R^2 and average alphas, rest-of-world models demonstrate superiority when considering the Gibbons et al. (1989) (GRS) statistic and Barillas et al. (2020) Sharpe ratio tests. I thus find that the conclusion on the superiority of local, regional, or rest-of-world factor models depends on the test metric.

Lastly, to determine an optimal factor asset pricing model, I follow the procedure by Swade et al. (2024) which iteratively builds a factor model intending to optimize either R^2, average alphas, or the GRS statistic. To maximize R^2 the optimal model consists of mostly local factors but concerning the other two metrics, the optimal model consists of local, regional, and rest-of-world factors. To dig deeper into this matter, I run the refined version of the Barillas and Shanken (2018) test (Choi et al., 2022) that assesses what subset of factors prices all other local, regional, or rest-of-world factors. I find that the optimal model mostly consists of regional and rest-of-world factors. Furthermore, I can show that the superiority of local asset pricing models can mostly be attributed to the local market factor.

I conclude that there are significant regional dependencies in factor returns that should be recognized when testing factor models internationally. The strength of these dependencies varies among factor types but generally separates countries along geographical and economic lines. In asset pricing tests, the performance ranking of factor models varies depending on the evaluation metric employed, with local models excelling in certain metrics while more global models outperform in others. However, local asset pricing factors are not over-represented in optimally constructed models and their superiority concerning certain performance metrics can partly be attributed to the local market factor.

The literature on asset pricing models encompasses empirical, theoretical, and predictive dimensions, with ongoing debates surrounding the efficacy of local, regional, and global factors.

As mentioned before, many empirical studies like those by Fama and French (2012, 2017), Griffin (2002), Hollstein (2022), and Hou et al. (2011) consistently provide support for local asset pricing models. However, these findings are not without contention, as highlighted by

Petzev et al. (2016), who track the evolution of local and global models over time, noting an increasing explanatory power of global models. Furthermore, Hau (2011) employs a natural experiment in a major MSCI index, suggesting a global rather than local pricing mechanism. In reconciling these empirical discrepancies, Brooks and Del Negro (2005) suggests advocating for regional factor models when pricing local assets. Theoretical literature predominantly favors global asset pricing models (e.g., Grauer et al., 1976; Solnik, 1974; Stulz, 1981), yet empirical evidence, as summarized above, often aligns with local models. A comprehensive review of the literature is provided in Karolyi and Stulz (2003). My work reunites these theoretical models with the prevailing empirical findings as I find that the answer to the best model most likely consists of local, regional, and global components.

Extending the debate beyond stocks, research by Ilmanen (1995) and Longstaff et al. (2011) explores the local vs. global asset pricing dynamics in bonds, while Brandt et al. (2006) delves into the implications for exchange rates. Another closely related literature stream is that of prediction as it is argued that stock characteristics can always be seen as the stock's factor loading to some latent factor (Kelly & Xiu, 2023, Lemma 1). Contrary to empirical asset pricing, where local models often prevail, for predictive models more global models seem to perform better. It is reported that the inclusion of U.S. data for international stock return prediction increases performance (Choi et al., 2022; Tobek & Hronec, 2021). Similarly, Rapach et al. (2013) show that lagged U.S. market returns can predict market returns of other countries. Hellum et al. (2023) quantifies the regionality in prediction and reports that 94% of return predictability follows global parameters. Finally, broader considerations such as home bias, flows, integration, and globalization, as reviewed by Lewis (2011) and Bekaert et al. (2016), underscore the multifaceted nature of asset pricing dynamics in an increasingly interconnected world.

The structure of the remaining article is as follows: Section 2 introduces the data. Section 3 detects what regional structure can be found in international factor returns and defines the regions that are used in later analyses. Section 4 contains the asset pricing analysis of the paper. It presents the results from spanning tests, examines the ability of local, regional, and rest-of-world factor models to explain asset returns, creates optimal factor models, and controls the results for the local market factor. Section 5 concludes and lays out the implications of my work. The appendix holds supplementary material such as additional tables and figures, a detailed description of the sorting variables, and technical details for some statistical tests.

2 Data

My data sample covers 35 stock markets around the world between January 1992 and December 2023. This sample of countries represents 84% of the World GDP in 2022 and is thus a representative sample for global factor analyses.[1] Return data for the U.S. market is sourced from CRSP while all other return data and accounting variables are gathered from Compustat. Following international asset pricing literature (i.e. Fama & French, 2012, 2017), returns are in USD based on Compustat exchange rates. For some parts of the paper, I use local currency returns but if not stated otherwise, returns are in USD. The risk-free rate is proxied by the 1-month Treasury rate. To minimize errors of Compustat returns, I winsorize returns at the 0.01% and 99.9% level cross-sectionally as in Jensen et al. (2023).[2] A comprehensive summary of the data is provided in Table A.1.

For the primary analysis, I employ the six-factor model by Fama and French (2018). This selection is motivated by two key reasons. First, to analyze the heterogeneity in regional dynamics of factors, it is appropriate to use a model with a relatively large number of factors. Second, this model has demonstrated superiority over the widely used five-factor model by Fama and French (2015) (Barillas et al., 2020; Fama & French, 2018). Though the main analysis will use the six-factor model, I will show that the use of other models such as the q-factor model by (Hou et al., 2015), the five-factor model by Fama and French (2015), the three-factor model by Fama and French (1993), as well as the replacement of the accruals based profitability factor with a cash profitability factor as proposed by Ball et al. (2016) will lead to similar results.

With regards to the factor construction, I closely follow Jensen et al. (2023). Each month, I sort stocks into tercile characteristics portfolios with breakpoints based on non-micro stocks.

[1]This calculation is based on global USD-converted GDP data from the Worldbank database.

[2]E.g., in January 1992 in Argentina when the country replaced its the Austral with the Peso at a rate of 10,000 to 1 returns of more than 100,000% are reported in the data for a few stocks . This is presumably due to an error in the conversion of exchange rates and should be corrected.

Micro-stocks are spread equally across those portfolios thereafter to make the portfolios more tradable. I require at least 5 stocks in each of the long and short legs. For each leg I compute the capped value weight return, meaning that I weight stocks by their market equity winsorized at the NYSE 80$^{\text{th}}$ percentile. Factor returns are a long-short zero investment strategy between the upper and lower portfolio. Though proposed by some of the original authors of the respective factor models I deviate from their methodology that uses double or triple sorts (e.g., Hou et al., 2015). I do this to ensure a large sample of countries to uncover regional dynamics and such sorting procedures result in multiple portfolios for which a large sample of stocks is required. However, I show later that my results also hold if factors are built following the traditional double sorting procedure by Fama and French (2017).[3]

The six-factor model is composed of a market factor, which is defined as the cross-sectional value-weighted excess return of stocks, and five long-short factors—a size factor, that sorts stocks on market capitalization, a value factor, that sorts stocks on book-to-market ratios, an investment factor, that sorts stocks on asset growth, a profitability factor, that sorts stocks on operating profits to book equity, and a momentum factor, that sorts stocks on their past 12-1-month return.[4]

3 Structure Detection

To justify regional asset pricing tests there should be some regional integration of factor returns. If they are regionally integrated, it can be expected that they also show some co-movement. Because of this reasoning, I examine the correlation structure of factor returns in the following paragraphs. This analysis helps answer the question of whether there is enough regional integration to justify regional testing and along what borders those regions should be defined.

Studies such as those by Rapach et al. (2013) examine the effects of market returns across different countries and find significant spillover effects. However, no research to date has investigated the international structure of factor returns. To ensure that results are not

[3] Due to the more data intensive double sorting procedure, the sample size decreases to 21 countries in this setting.

[4] A detailed description of the sorting variables used for the long-short portfolios can be found in Appendix B.

influenced by any residual market exposure, I first regress all factor returns on their respective market returns as

$$f_\tau = \alpha_\tau + \beta_\tau r_m + \epsilon_\tau \tag{1}$$

for each factor, τ, and only use the residuals to perform the following analyses. In the above equation, f_τ is a vector of factor returns and r_m is a vector of respective market returns. This cleaning approach is particularly important to isolate the pure factor exposure since I find that some of the factors exhibit modest correlations with the market factor (see Figure A.1).

Another cause for spurious correlation could be the influence of currency returns as all returns are converted into USD. To assess whether the currency conversion affects the results, I repeated the below analysis for factor returns in local currency returns. I find that the results are identical to those outlined below, implying that the structure I find is not caused by currency effects.[5]

3.1 Individual Factor Analysis

To get a first impression of whether there is co-movement in international factor returns, I calculate the complete Pearson correlation matrix, $\mathbb{C}$, for all country-wise factor returns separately for each factor as $\mathbb{C} = S^{-1/2}\Sigma S^{-1/2}$, where Σ is the covariance matrix of factor returns and S is the diagonal matrix of Σ. Σ is calculated as $\Sigma = F^\top F$, where F is a matrix of factor returns for one factor type (i.e., value),[6] with T rows and N columns, where T is the total number of months and N is the total number of countries in my sample.

In total, I calculate six correlation matrices[7] that are illustrated in Figure 1. The countries are

[5]The results of the iPCA Analysis with local currency returns among other methodological robustness checks are illustrated in Figure A.5.

[6]Following the earlier notation all variables would require a subscript τ that identifies the respective factor types (i.e., F would be F_τ). For ease of readability this subscript is dropped in this section.

[7]Due to the heterogeneous availability of factor data among countries, the number of observations with which the correlation matrices can be calculated is not the same for each factor. The number of complete observations per factor is reported in Table A.2. To ensure better comparability with regards to the number of observations, the analysis has been redone using pairwise complete correlation matrices for which the results are almost identical and thus not reported here.

ordered according to their respective continents and we can see that for market returns, there is co-movement within continents. The value and momentum factors show similar patterns which cannot be observed for the profitability factor. International size and investment factor returns also show some co-movement but not as strong as that for value and momentum. This already provides evidence that factor returns differ concerning their strength of regional integration.

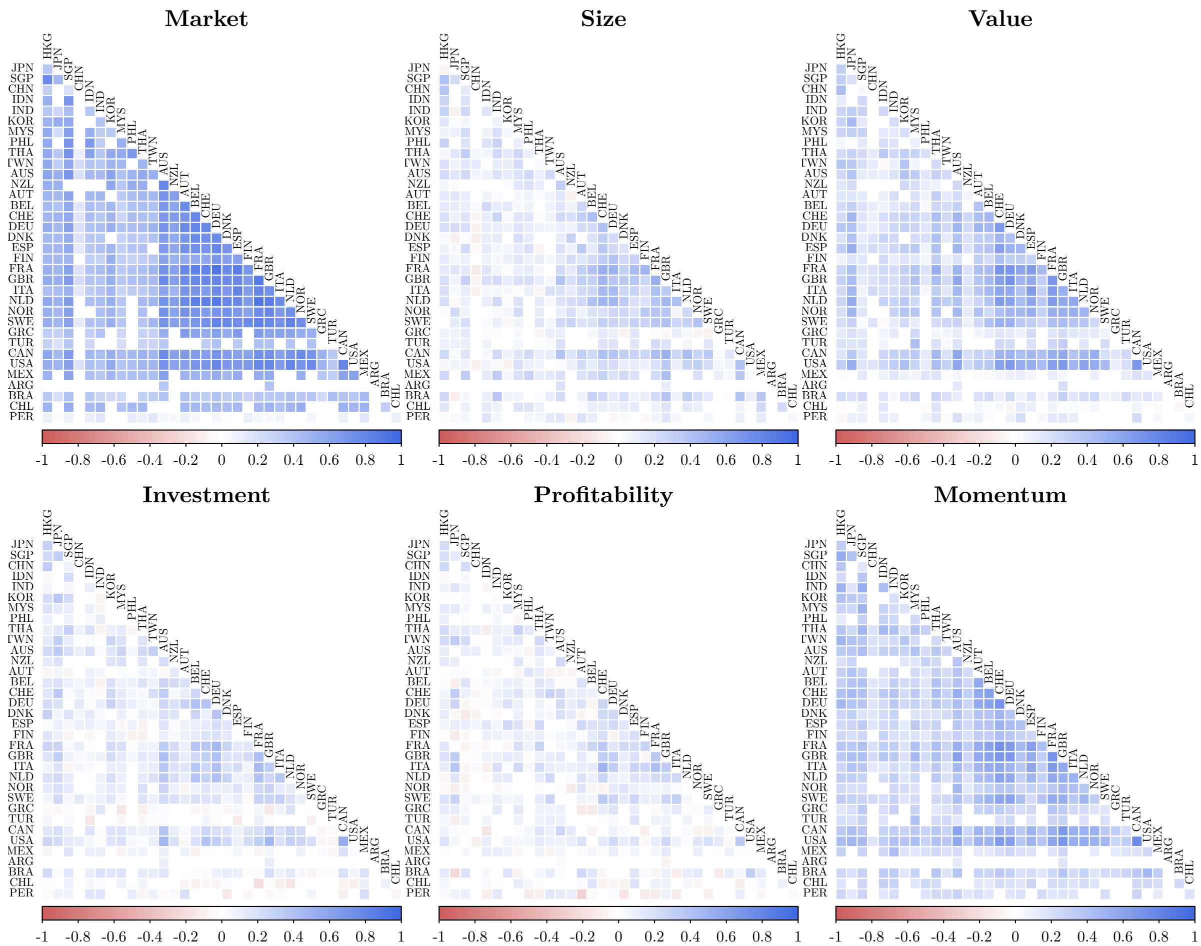

Figure 1: Correlation plots of global factor returns

The figure illustrates correlation matrices for monthly international returns in USD of the Fama and French (2018) six-factor model between 1992 and 2023 that were cleaned for spurious market exposure. The matrix for each factor is plotted separately.

While the correlation matrices grant interesting insights into the co-movement, they cannot be used to assess the similarity or distance between the factor returns. To overcome this hurdle, I perform principal component analysis (PCA) using an eigendecomposition of each of the correlation matrices for the international factor returns. This procedure provides axes among which the data varies the most, which can be used for determining distances between countries. Furthermore, I use this dimensionality reduction algorithm because it is helpful to extract spurious correlation from the signal. This is necessary as it is reported that in finance the signal-to-noise ratio is very low (e.g., Kelly et al., 2019).

The correlation matrix of returns, $\mathbb{C}$, can be decomposed as,

$$\mathbb{C} = PDP', \qquad \text{with} \quad D = \operatorname{diag}(d_1, d_2, \ldots, d_N) \tag{2}$$

where $P = (v_1, \ldots, v_N)$ is the matrix of eigenvectors v_i and D is a diagonal matrix of eigenvalues d_i ordered in decreasing magnitude. From the definition of the eigenvector and eigenvalue, we know that $\mathbb{C}v_i = d_i v_i$ and that $d_i / \sum d_i$ is the quantity of total variance explained by the corresponding v_i. The eigenvectors that correspond to the largest eigenvalues thus represent the axis among which the data is most spread out.[8] The literature on PCA often describes the underlying data in terms of samples and features. For my application, I treat each time stamp as a sample of international factor data for which each country-wise return represents a feature.

To get an impression among what axis the factor return data varies the most, I perform the eigendecomposition and plot v_1 and v_2 in Figure 2. The eigenvectors are examined because they represent scaled versions of the principal components (PCs) (Tang & Allen, 2021). Thus, these axes can be interpreted as the distance between factor returns. Since I use the correlation matrix for the decomposition, which effectively normalizes returns, their mean and volatility do not affect the results.

[8] It is important to note that using the correlation matrix ensures that the eigenvectors represent the axes along which the data is most spread out jointly. This approach effectively normalizes the returns, which is desired for the undertaking of this study. Without normalization, if a factor return is significantly more volatile in one country compared to others, the decomposition might produce an eigenvector dominated by that country's variance, thereby isolating it from the rest. Since the focus of this study is to determine joint variation across countries, using the correlation matrix eliminates the impact of such variance differences, which are not of primary interest in this analysis.

We can see that the first two eigenvectors group countries with similar development status and geographic region. This pattern is relatively similar for all factors. However, the strength of this clustering is very different. Consistent with the observations of the correlation matrices, the variance explained by the first two PCs is highest for the market factor at close to 60% and lowest for the profitability factor at around 18%. This suggests that although the regional dynamics are relatively consistent across different factors, their magnitudes vary significantly.

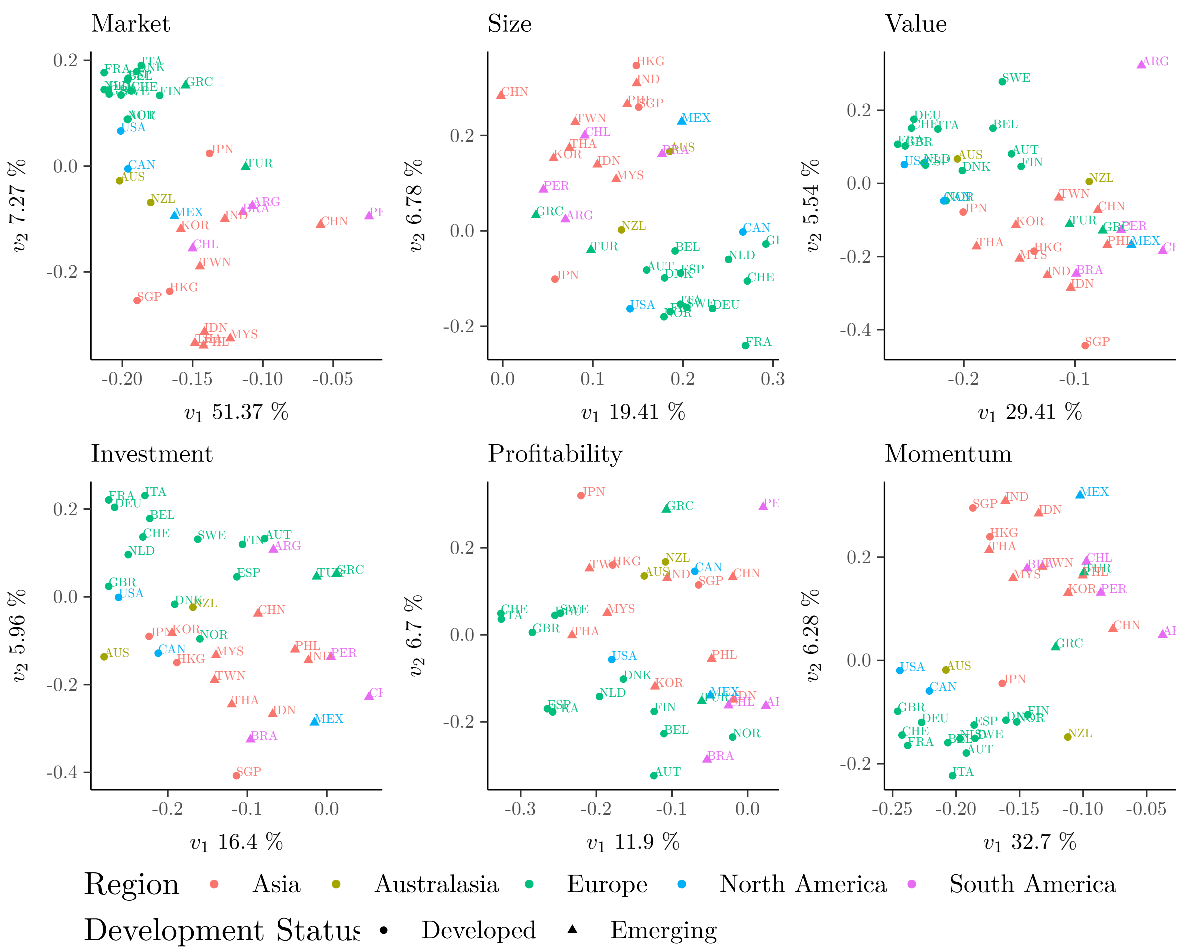

Figure 2: Principal component analysis of international factor returns

The figure holds the first two eigenvectors, v_1 and v_2 of correlation matrices from international Fama and French (2018) six-factor model returns in USD between 1992 and 2023 that were cleaned for market exposure. The eigenvectors are plotted for each factor separately. The percentages on the axes describe the portion of explained variance by the respective eigenvector. Development statuses are sourced from MSCI.

3.2 Joint Analysis

The following section lays out an integrated analysis of all factor returns from which fully data-driven factor regions are defined. An integrated analysis is possible because the above evidence suggests that (if they exist) the regional dynamics are relatively similar for different factors. The extraction of regions from the factor data will involve five steps: first, jointly decomposing the factor data; second, determining the number of dimensions to retain; third, assessing whether the retained data forms clusters; fourth, defining the optimal number of clusters; and fifth, retrieving the cluster allocation for each country. The following paragraphs will outline each of the five steps within which I am employing a battery of different statistical metrics. To not dilute the text with technical details, I shift the technicalities to Appendix C. Here, I will provide intuition for what the statistical measures do and interpret the results.

Step 1: Joint Decomposition: For the first step of the analysis, I use the integrated PCA (iPCA) algorithm by Tang and Allen (2021). This method allows to decompose multiple matrices simultaneously and is a generalized form of PCA. It is particularly useful for the task at hand for two reasons. First, iPCA aims to extract dominant joint patterns which are common to multiple data sets, not necessarily the variance-maximizing patterns since they might be specific to one data set. Second, it treats each of the matrices as a standalone sample of observations, which comes in handy as the number of complete observations varies among factors.[9]

The first two eigenvectors that result from the iPCA analysis are plotted in Figure 5. As expected, the emerging pattern looks very similar to that for individual factor returns in Figure 2. Again, countries are grouped along continents and development status. It is worth noting that the variance explained by the first two principal components is now much lower than in the setting for individual factors. This is expected because, unlike traditional PCA, the iPCA algorithm does not focus solely on extracting variance-maximizing patterns. Instead, it also aims to find joint patterns in the data, resulting in lower explained total

[9]Due to the heterogeneous availability of factor data among countries, the number of observations with which the correlation matrices can be calculated varies for each factor. The number of complete observations per factor is reported in Table A.2. Furthermore, to ensure that the results are not excessively influenced by the market factor—which has more complete observations available than the other factors—I run the iPCA algorithm for the Fama and French (2018) six-factor model, excluding the market factor. The results for this analysis among other methodological robustness checks are illustrated in Figure A.5.

variance. The results are insensitive to the method as other integration algorithms such as the Multiple Factor Analysis (MFA) (Abdi et al., 2013; Escofier & Pagès, 1994) yield very similar results.[10]

Step 2: Dimension Reduction: The second step of the analysis involves determining the appropriate number of eigenvectors to retain. Initially, for the sake of easier graphical interpretation, the first two eigenvectors were examined in the eigenvector plots above. However, the optimal number of dimensions to retain may differ from this number. To extract the number of informative dimensions, I utilize the parallel analysis method by Horn (1965) that adjusts eigenvectors for sampling errors using bootstrapping. Effectively it adjusts the eigenvalues by randomly generated noise. After this adjustment, eigenvectors whose eigenvalue is larger than 1 are deemed to carry information about the underlying data.

I perform this adjustment with 10,000 bootstrap iterations for which the results are visualized in Figure A.2. The first two eigenvalues comply with the criterion of being larger than 1 and are thus retained for the following analyses.

Step 3: Assess Clustering Tendency: After having retained the informative part of the eigendecomposition, it has to be determined whether the observations are clustered. For this task, I use the clustering statistic by Hopkins (1954). The statistic compares the nearest-neighbor distribution of randomly generated data points in the observed span to that for randomly selected data points. If the data was randomly distributed, the statistic should be about 0.5, and a value of more than 0.7 is deemed to indicate clustering (Banerjee & Davé, 2004).

I estimate the statistic using 10.000 bootstrapped draws for two different settings[11]. Both setups lead to similar results (0.76 and 0.73), indicating that the data is clustered. For one of the two settings, the result is statistically different from 0.5 at 10% confidence.

Step 4: Optimal Number of Clusters: Knowing that the data is clustered, the optimal number of clusters has to be determined. For this task, I use three different metrics. First, I use the within sum of squares (WSS) which calculates the average distance between points

[10]The results for the MFA algorithm among other methodological robustness checks are illustrated in Figure A.5.

[11]The exact setup of these settings can be found in Appendix C.

within a cluster. The optimum number of clusters is the point where the WSS cannot be improved strongly for one more cluster (elbow method). Second, I use the silhouette method by Rousseeuw (1987) which takes the distance of each data point to all other points within its cluster and compares it to the average distance of this point to all other points outside its cluster. The optimal number of clusters maximizes the average silhouette value of all data points. Third, I use the gap statistic by Tibshirani et al. (2001). In a nutshell, the gap statistic compares the average intra-cluster variation of the actual clusters to the intra-cluster variation of uniformly randomly generated data. This statistic is maximized for the optimal number of clusters.

I calculate the above three metrics for different numbers of clusters between one and ten using the hierarchical agglomerative clustering method (Murtagh & Legendre, 2014) with the linkage criterion of Ward (1963).[12] The results are illustrated in Figure 3. It can be seen that for more than three clusters, the WSS can no longer be improved by much. The silhouette and gap statistics show a similar result as they are maximized for three clusters. Thus, the optimum number of clusters is three.

Step 5: Define Final Clusters: The concrete cluster allocation of each cluster can directly be retrieved from the hierarchical cluster analysis in step 4 with three clusters. The resulting dendrogram of this specification is illustrated in Figure 4 within which I highlight three regional clusters that I call ”established”, ”emerged” and ”developing”.[13] The established region comprises developed European countries as well as Canada, the U.S., and Australia. It thus represents developed Western countries. The emerged region includes the most developed Asian countries, such as Japan, Korea, and China. Lastly, the developing region comprises developing countries from all continents, such as Turkey, Argentina, and India. The classification of the hierarchical clusters for the results of the iPCA analysis is illustrated in Figure 5. It can be seen that three dense clusters were created. Also, consistent with the picture from the visual inspection of the first two eigenvectors, the clustering of the countries

[12]To validate the results for the choice of clustering algorithm, I calculate the above summary statistics again using the k-means clustering method and the Partitioning Around Medoids (PAM) algorithm (Kaufman & Rousseeuw, 1990) instead of hierarchical clustering. The results are illustrated in Figure A.3 and Figure A.4 and look almost identical to those presented here.

[13]Although the naming of these three data-driven clusters is heavily inspired by the two development statuses, ”developed” and ”emerging,” they should not be confused with them.

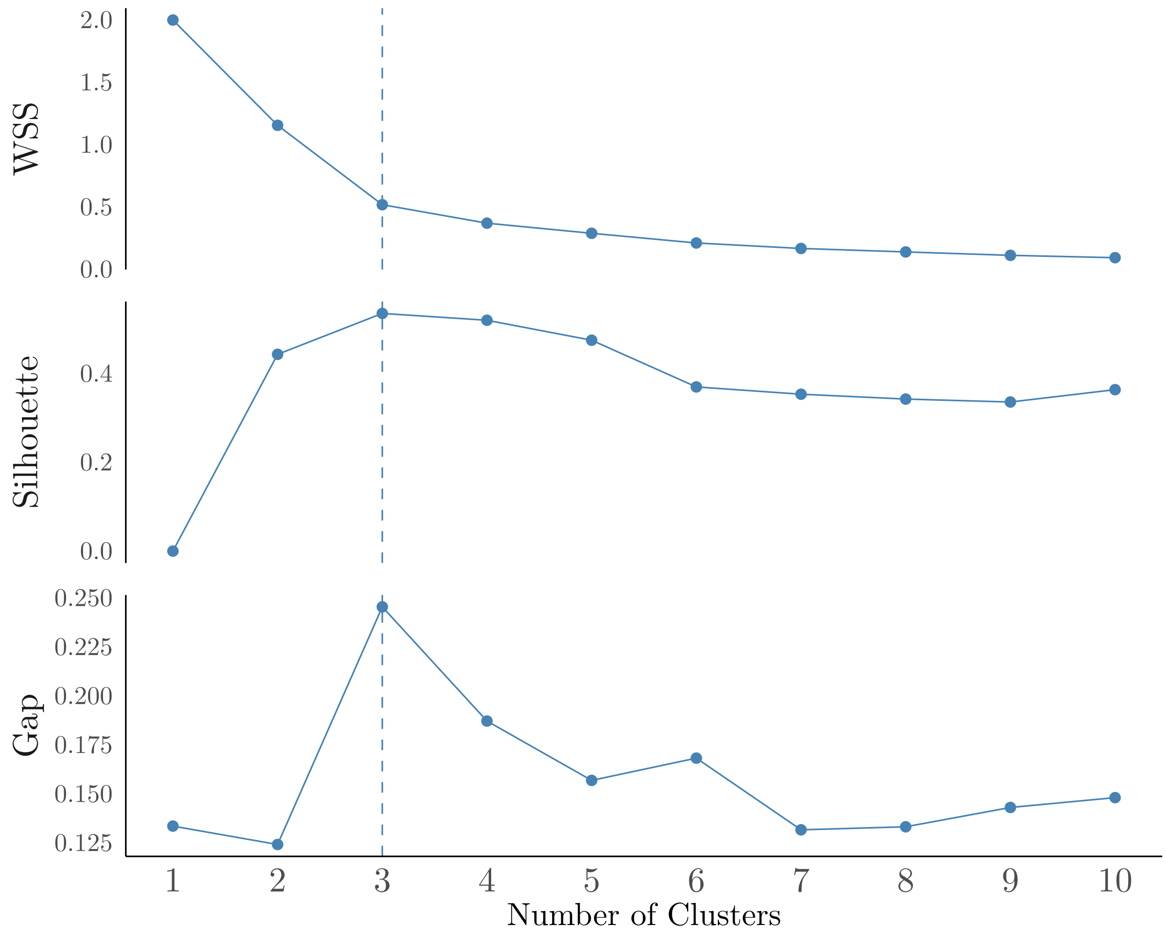

Figure 3: Clustering statistics using the hierarchical clustering

The figure holds three clustering measures for the first two eigenvectors, of correlation matrices from international Fama and French (2018) six-factor model returns in USD between 1992 and 2023 that were cleaned for market exposure. The eigenvectors are calculated jointly using the iPCA procedure by Tang and Allen (2021). "WSS" is the average distance between points within a cluster. "Silhouette" is the silhouette statistic by Rousseeuw (1987) which compares the distance of each data point within the cluster to the average distance to all other points outside the cluster. "Gap" is the gap statistic by Tibshirani et al. (2001) which compares the average intra-cluster variation to the intra-cluster variation of uniformly randomly generated data. The x-axis holds the number of respective clusters for which the measures were calculated. The clusters are generated using hierarchical agglomerative clustering (Murtagh & Legendre, 2014) with the linkage criterion of Ward (1963).

follows along geographic lines and development status.

A robustness check that the resulting clusters do not result from my modeling choices can be found in Figure A.5 of the appendix. The figure plots the first two eigenvectors for different modeling settings (local currency, MFA decomposition, exclusion of market factor, factor construction following the traditional factor formation methodology by Fama and French (2017), and other factor models). Despite flipped signs for some of the eigenvectors, the clustering is very similar to the setting employed here. It is worth mentioning that in none of the scenarios explored, I find compelling reasons to consider Japan as an outlier warranting the creation of its own distinct region, a practice commonly employed in other studies such as Fama and French (2012) and Hollstein (2022). However, unlike what one might expect based on economic proximity, Japan is not part of the established cluster (despite being close) but rather falls into the emerged cluster to which it shows a high geographical proximity. The only country that nearly qualifies as a single-country region is Argentina. Its distance to any other country is very large which causes it to be merged very late into a cluster in the dendrogram below.

Lastly, to determine the strength of regional integration for the different factors I calculate the average correlation for the factor returns in Table 1. It can be seen that the established region shows the largest average correlation across factor returns. This is consistent with the results of the clustering analysis as in this region many countries fall very closely together in Figure 5. The emerged region shows less co-movement and the developing region has very little to no integration for most long-short factor returns. Consistent with the PCA analysis, the factors for which most variance is explained by the first two PCs also show the highest average correlation. If we take market returns as a benchmark, the regional integration for the most integrated factors (like value and momentum) can reach up to 68% of the market return's integration.

4 Asset Pricing

To test whether asset pricing is more local, regional, or global, I create three different factors for each factor class in each country. I define a local factor in country c as f_c^L following the procedure outlined in Section 2. Note, that I suppress the time subscripts to make the notation cleaner. Next, I define a regional factor as the market capitalization-weighted

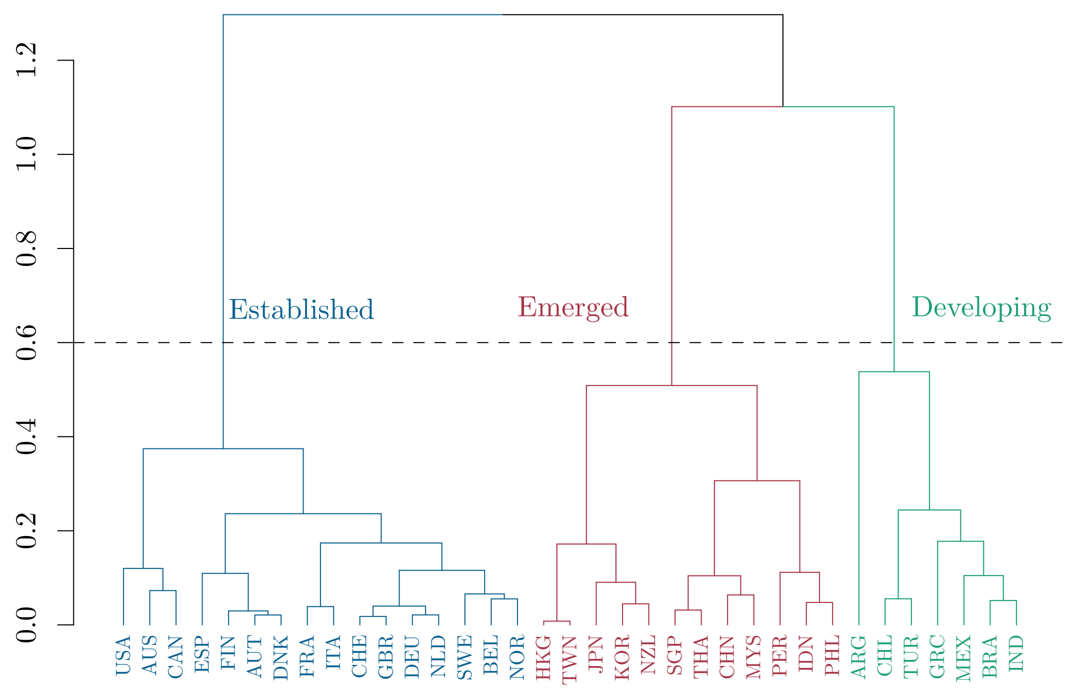

Figure 4: Dendrogram for international factor returns

The figure holds a dendrogram resulting from a hierarchical agglomerative clustering (Murtagh & Legendre, 2014) with the linkage criterion of Ward (1963). The underlying data are the first two eigenvectors, of correlation matrices from international Fama and French (2018) six-factor model returns in USD between 1992 and 2023 that were cleaned for market exposure. The eigenvectors are calculated jointly using the iPCA procedure by Tang and Allen (2021).

average factor return for all other countries that are in the same region as c as

$$f_c^R = \sum \frac{m_k}{(m_c^R - m_c)} f_k^L \qquad \forall k \in R_c, k \neq c, \tag{3}$$

where R_c is the region that includes c, m_i is the market capitalization of country i, and m_c^R is the total market cap of countries in that region.[14] Lastly, I define a rest-of-world factor as the market capitalization-weighted average factor return for all countries outside of R_c as

[14]Note the slight abuse of notation. Whenever I refer to the region of country c, I denote it with R_c. However, whenever I refer to some other variable that is determined by R_c I put the c in the subscript of the main variable to make the notation cleaner. For example, I denote the regional factor return in R_c as f_c^R.

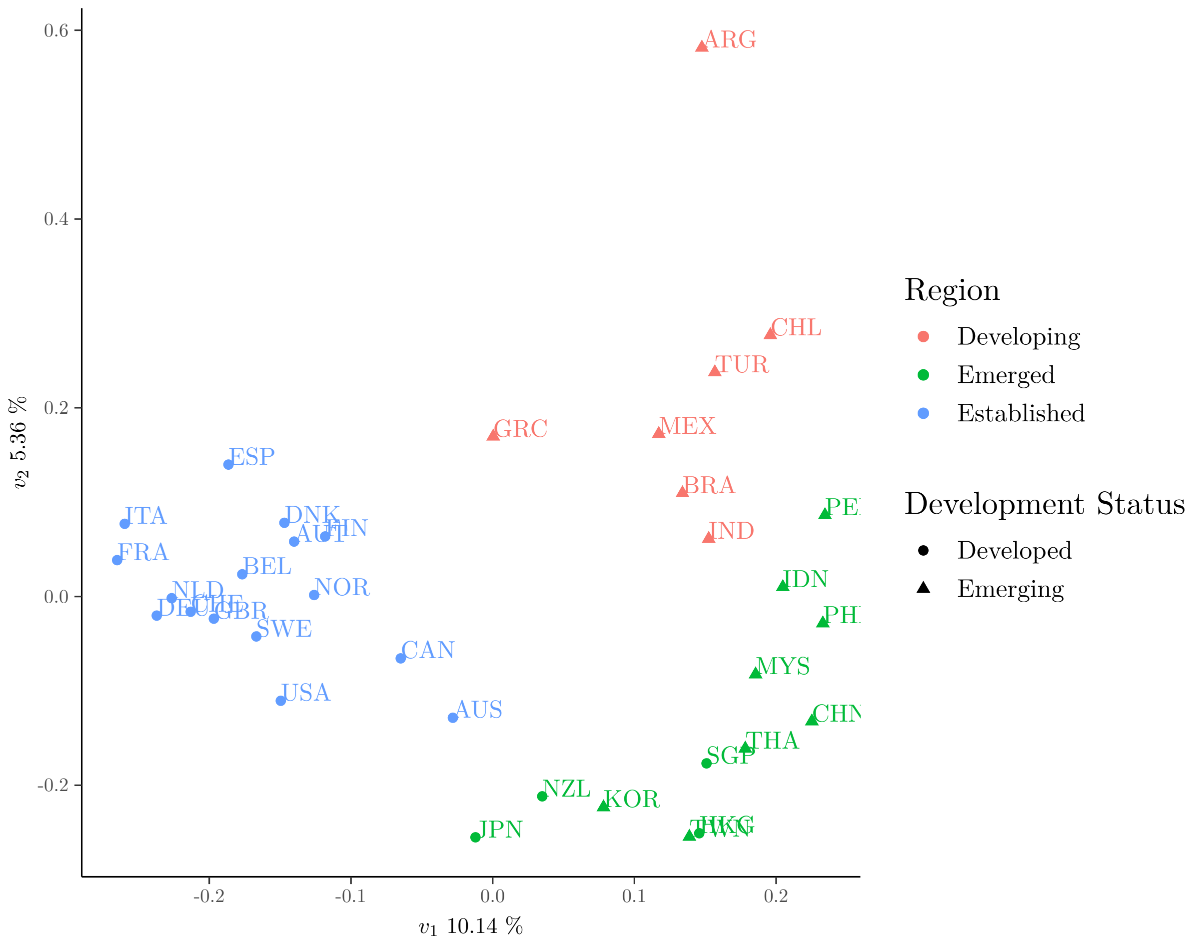

Figure 5: Principal component analysis of global factor returns

The figure holds the first two eigenvectors, v_1 and v_2, of correlation matrices from international Fama and French (2018) six-factor model returns in USD between 1992 and 2023 that were cleaned for market exposure. The eigenvectors are calculated jointly using the iPCA procedure by Tang and Allen (2021). The percentages on the axes describe the portion of explained variance by the respective eigenvector. Development statuses are sourced from MSCI. The regions are assessed based on the results of a hierarchical agglomerative clustering (Murtagh & Legendre, 2014) with the linkage criterion of Ward (1963) of the data using three clusters.

$$f_c^W = \sum \frac{m_j}{(M - m_c^R)} f_j^L \qquad \forall j \notin R_c, \tag{4}$$

where M is the total market capitalization of all countries in my sample. I set the minimum number of countries from which f_c^W and f_c^R can be built to two. Note, that f_c^R is unique for each country as c is always excluded from its respective regional factor return while

Table 1: Average correlations of factor returns within regions

The table holds the within-region average correlations of country-wise Fama and French (2018) six-factor model returns in USD between 1992 and 2023. The regions are assessed using hierarchical agglomerative clustering (Murtagh & Legendre, 2014) with Ward's linkage criterion (Ward, 1963) on the first two eigenvectors of international factor correlation matrices that were jointly extracted using the iPCA procedure by Tang and Allen (2021).

	Developing	**Emerged**	**Established**
Market	0.35	0.40	0.73
Size	0.14	0.10	0.26
Value	0.06	0.09	0.40
Investment	0.13	0.15	0.25
Profitability	0.10	0.13	0.17
Momentum	0.19	0.26	0.50
Total	0.16	0.19	0.38

f_c^W is the same for all c from the same region. Following Asness et al. (2015), Ehsani and Linnainmaa (2022), Gupta and Kelly (2019), and Jensen et al. (2023), among many others, I scale all factor returns to have a yearly volatility of 10%.[15] My definitions for regional and global factors differ from current literature (e.g., Fama & French, 2012; Hollstein, 2022) that does not exclude c (constituents of the region) from its region (the global model) and uses sorting portfolios from the regional/global pool instead of aggregating country-level factors to regional/global factors. The motivation for this differentiation is twofold. I perform the exclusion of c (the respective region) to not contaminate regional (global) factors with local (regional) information. Also, I aggregate local factor portfolios so all countries are represented in the respective regional/global portfolio. This is not guaranteed if such portfolios were created from sorting stocks in the regional/global pool of assets.[16] Nevertheless, I show in Table A.6 and Figure A.8 that the main results of my asset pricing analysis are robust to

[15]The reason for this is twofold. First, as argued by Jensen et al. (2023), ensuring that factors have the same variance provides a prior that factors are similar in terms of their information ratio (i.e., appraisal ratio). This is crucial because I later conduct Bayesian testing and investigate maximum Sharpe ratio portfolios, where it is beneficial for each factor to have a similar risk contribution. Second, I perform spanning regressions and examine the coefficients relative to other factors. For ease of comparability, it is essential that the factors have the same variance.

[16]E.g., in case of a regional value portfolio, if stocks in one country have very extreme book-to-market ratios, the regional portfolio will be mostly composed of stocks from that country. This is not an unlikely scenario as it is shown that e.g. book-to-market ratios are fundamentally different between European and U.S. banks (Simoens & Vennet, 2021) and that large differences in valuation ratios can be caused by different accounting regulations among countries (Arce & Mora, 2002)

the choice of the portfolio construction. To make my results more comparable to those of the existing literature, I use the raw factor returns instead of the residuals from (1) for the following parts of the analysis.

4.1 Spanning Tests

To begin, I assess the strength of the regional dependencies for factor returns. In doing so, I test whether the regional factor returns encompass their local counterparts. This is accomplished by conducting the following spanning regression for each country, c, and factor, τ, against regional factor returns from all three regions:

$$f_{c,\tau}^{L} = \alpha_{c,\tau} + \sum \beta_{c,\tau} f_{\tau}^{R} + \varepsilon_{c,\tau} \tag{5}$$

If the regional dependencies are strong, the local factor should load heavily on its regional factor. The other regional factors can be interpreted as a control terms. All $\beta_{c,\tau}$ and $\alpha_{c,\tau}$ that are significant at 5% are plotted in Figure 6. We can see that those factors for which the structure analysis shows stronger regional integration (e.g., value and momentum) load more heavily on their respective regional factors. We further see that this loading is stronger in the established region. For the emerged and developing regions the pattern cannot be observed. The results provide evidence that in the established region the information of local factors is well included in their regional counterparts.

To make my results more comparable to existing work, I repeat the spanning analysis but only include the factor of the respective region on the right side of (5). The equation thus simplifies to

$$f_{c,\tau}^{L} = \alpha_{c,\tau} + \beta_{c,\tau} f_{c,\tau}^{R} + \varepsilon_{c,\tau}. \tag{6}$$

The relative number of insignificant $\alpha_{c,\tau}$ is reported in Table 2. We can see that even if local factors are only regressed on their regional counterpart, most local factors are spanned by their regional counterparts. This conclusion contradicts that by Hollstein (2022) who argues

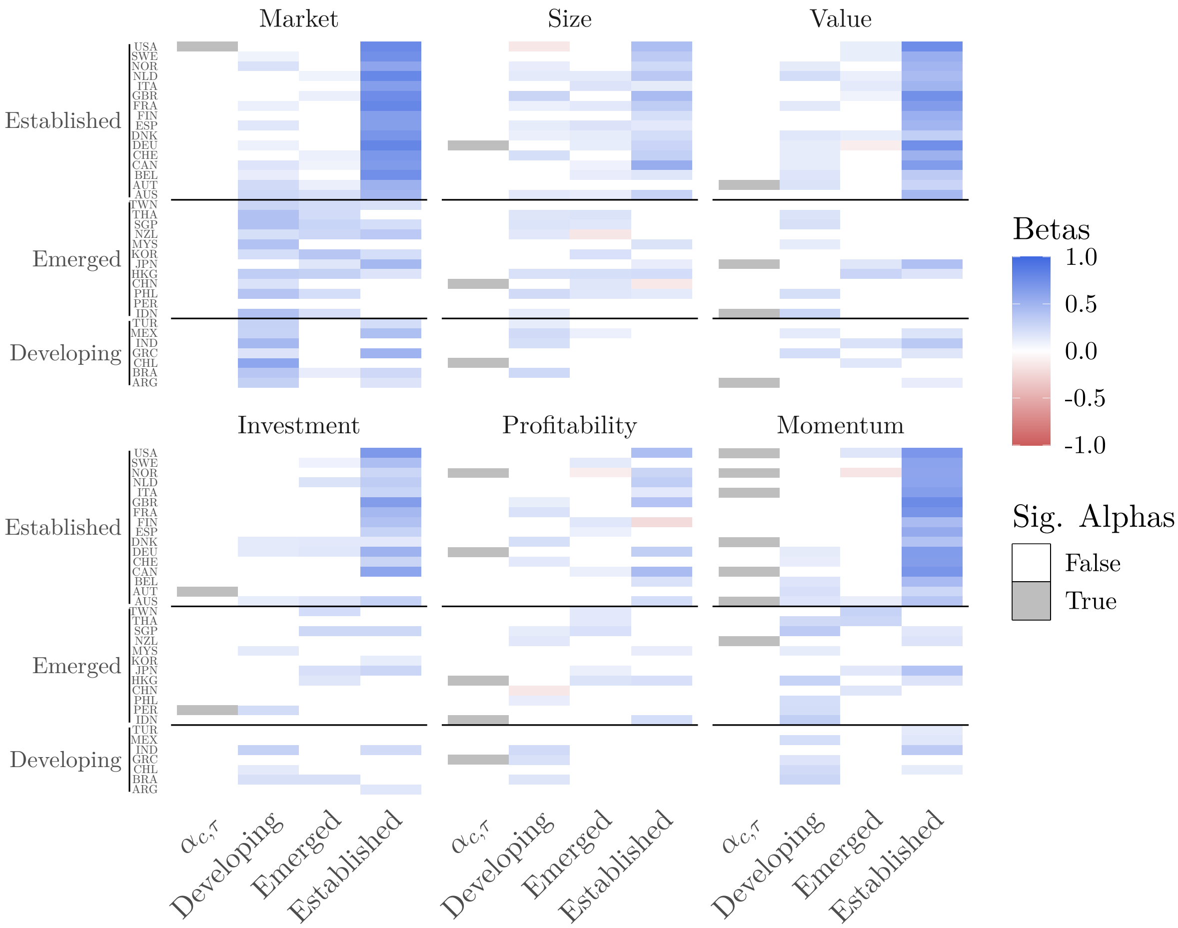

Figure 6: Spanning regressions of regional factor returns

The figure holds the results of spanning regressions for international Fama and French (2018) six-factor model returns in USD between 1992 and 2023. The regression setup is $f^L_{c,\tau} = \alpha_{c,\tau} + \sum \beta_{c,\tau} f^R_\tau + \varepsilon_{c,\tau}$, for each country, c, and factor, τ where $f^L_{c,\tau}$ is the local factor return and f^R_τ are regional factor returns from the developed, emerged, and established region. A local factor consists of sorted long-short stock portfolios in the respective country. A regional factor is a value-weighted average of local factors in the respective region, excluding the respective country. The regression equation is fitted with GMM. Confidence intervals are calculated using Newey and West (1987) adjusted standard errors using 6 lags. The plot holds $\beta_{c,\tau}$ and $\alpha_{c,\tau}$ coefficients that are significant at 5% confidence.

for weak spanning results. The reasons for this contradiction are twofold. First, Hollstein (2022) tests whether complete asset pricing models, that contain multiple factors, can be spanned by their regional counterparts. Such a joint test is harder to pass and thus leads to more rejections. Furthermore, Hollstein (2022) concludes weak spanning given his finding of sizable alpha intercepts. When looking at the statistical significance reported in his study, we can see that though sizable, the hypothesis that these alphas are different from zero can not be rejected in most cases, similar to the results reported here.

Table 2: Averages of insignificant $\alpha_{c,\tau}$ in regional spanning test

The table holds the results of spanning regressions for international Fama and French (2018) six-factor model returns in USD between 1992 and 2023. The regression setup is $f^L_{c,\tau} = \alpha_{c,\tau} + \beta_{c,\tau} f^R_{c,\tau} + \varepsilon_{c,\tau}$, for each country, c, and factor, τ where $f^L_{c,\tau}$ is the local factor return and $f^R_{c,\tau}$ is the respective region of country c. A local factor consists of sorted long-short stock portfolios in the respective country. A regional factor is a value-weighted average of local factors in the respective region, excluding the respective country. The regression equation is fitted with GMM. Confidence intervals are calculated using Newey and West (1987) adjusted standard errors using 6 lags. The table holds the relative number of $\alpha_{c,\tau}$ coefficients that are insignificant at 5% confidence in each region and in total.

Factor	Develop-ing	Emerged	Estab-lished	**Total**
Market	100%	92%	94%	**94%**
Size	86%	92%	100%	**94%**
Value	86%	75%	94%	**86%**
Investment	86%	92%	94%	**91%**
Profitability	86%	83%	81%	**83%**
Momentum	100%	92%	56%	**77%**
Total	90%	88%	86%	**88%**

4.2 Pricing Test Assets

To test how well factors perform with regards to their asset pricing ability I assess how well local, regional, and rest-of-world factors can price test assets for which I use GICS 10 portfolios. This decision aims to include as many countries as possible due to the heterogeneous availability of data. Although data-rich countries like the U.S. would allow for more test assets, I use the same number of test assets across countries to ensure comparability of results.[17] The asset pricing tests are conducted by regressing the excess return of industry portfolio i on that of local, global, and regional factor models as

$$r_{c,i} = \alpha_{c,i} + \sum \beta_{c,i} f_c^{L/R/W} + \varepsilon_{c,i}, \tag{7}$$

where $f_c^{L/R/W}$ contains the returns of a local, regional or rest-of-world factor model. I

[17] For robustness, results for tests using size-value sorted portfolios as test assets can be found in the Appendix, Table A.3.

require at least 120 months of observations for an industry portfolio to be included in the analysis. Furthermore, I use the same number of observations for each factor model to ensure comparability between models. I fit the above regression using the generalized method of moments (GMM) and evaluate statistical significance using Newey and West (1987) adjusted standard errors with six lags.

Following Hollstein (2022), I use R^2 of the above regression as well as the average size of the regression alphas, $\bar{\alpha}_c$, to evaluate the goodness of the different factor models. Additionally, I include two more measures that are based on recent developments in the literature that propose to use Sharpe ratios to evaluate factor models (Barillas et al., 2020; Barillas & Shanken, 2017; Fama & French, 2018). First, I employ the Gibbons et al. (1989) (GRS) statistic. This statistic jointly evaluates whether all $\alpha_{c,i} = 0$. It can be calculated as

$$F_{GRS} = \frac{T(T-I-K)}{I(T-K-1)} \frac{\alpha^T \Sigma^{-1} \alpha}{(1 + Sh^2(f))}, \tag{8}$$

where T is the number of observations, I is the number of test assets, K is the number of factors, $Sh^2(f)$ is the maximum squared Sharpe ratio obtainable by the tested factors, α is a vector holding all $\alpha_{c,i}$, and Σ is the covariance matrix of the regression residuals $\varepsilon_{c,i}$. The statistic follows is distributed as $F_{GRS} \sim F(I, T-I-K)$. Note, that this statistic heavily depends on the ratio between $\alpha^T \Sigma^{-1} \alpha$ and $Sh^2(f)$. Thus, a better F_{GRS} can be achieved by a factor model that minimizes $\alpha_{c,i}$ and can obtain a high $Sh^2(f)$. Lastly, I use the squared Sharpe ratio criterion by Barillas and Shanken (2017) who build on the argument by Gibbons et al. (1989) that

$$\alpha^T \Sigma^{-1} \alpha = Sh^2(f, A_c) - Sh^2(f), \tag{9}$$

where A_c is the matrix of test asset returns $r_{c,i}$. When A_c includes all possible factors, $Sh^2(f, A_c) = Sh^2(A_c)$ and minimizing $\alpha^T \Sigma^{-1} \alpha$ is equivalent to maximizing $Sh^2(f)$ which can be directly taken to evaluate asset pricing models and thus no test assets are required.

To evaluate whether the $Sh^2(f)$ values between competing models are statistically different, I employ the closed-form test statistic proposed by Barillas et al. (2020).

In theory, both F_{GRS} and $Sh^2(f)$ should rank factor models similarly, but their rankings can differ in practice. For example, consider two nested asset pricing models. According to $Sh^2(f)$, the model with more factors will always perform at least as well as the one with fewer factors, because the weight of the additional factor in the tangency portfolio can always be set to zero. However, if the additional factor has a negative slope coefficient, the average $\alpha_{c,i}$ increases for the model with more factors, potentially causing F_{GRS} to favor the model with fewer factors. Although it is shown that differences in rankings between the two measures are rare (Fama & French, 2018), both are included here for completeness.[18]

The results for R^2, $\bar{\alpha}_c$, and F_{GRS} are reported in Table 3. The results for R^2 and $\bar{\alpha}_c$ are consistent with the findings of Hollstein (2022). Local models perform better than regional and rest-of-world models as they have the highest R^2 and the lowest $\bar{\alpha}_c$. Also, we see that for the established region, the regional model has a higher R^2 than the rest-of-world model, whereas for the other regions, the two models perform more alike. For the F_{GRS} statistic the picture changes. Now the local model is the worst in terms of average F_{GRS} and the rest-of-world model performs the best. This is likely because more international factor models have higher $Sh^2(f)$, thus lowering the F_{GRS}. This intuition is supported by the results plotted in Figure 7 where I report a comparison of $Sh^2(f)$. We can see that regional and rest-of-world models achieve significantly higher $Sh^2(f)$ than local models. When comparing $Sh^2(f)$ between regional and rest-of-world models, we see that the rest-of-world model has higher $Sh^2(f)$ in the established region whereas for other regions they are fairly similar.

Overall, the asset pricing tests show that the conclusion on what factor model performs best heavily depends on the metric of choice. Consistent with Hollstein (2022), we would favor a local model if we look at R^2 and $\bar{\alpha}_c$. However, if we look at F_{GRS} or $Sh^2(f)$ we would favor more global models. The results are robust to using 3×3 sorted size-value portfolios as alternative test asset[19] for which the results are stored in Table A.3. Also, the results carry

[18]For a more detailed discussion on the differences between the two measures, see Barillas et al. (2020), Barillas and Shanken (2017), and Fama and French (2018).

[19]I acknowledge that the standard setting is to use 5×5 portfolios. The choice to reduce the total number of portfolios from 25 to 9 is based on the desire to include as many countries as possible in the analysis because

Table 3: Asset pricing tests local vs. regional vs. rest-of-world factor model with GICS portfolios

The table holds the results of asset pricing regressions for local, regional, and rest-of-world Fama and French (2018) six-factor model returns in USD between 1992 and 2023. The regression setup is $r_{c,i} = \alpha_{c,i} + \sum \beta_{c,i} f_c^{L/R/W} + \varepsilon_{c,i}$, for each country, c. $r_{c,i}$ are return of industry portfolio i for GICS 10 portfolios. $f_c^{L/R/W}$ contains the returns of a local, regional or rest-of-world factor model. The local model consists of sorted long-short stock portfolios in the respective country. The regional model is a value-weighted average of local factor models in the respective region, excluding the respective country. The rest-of-world model is a value-weighted average of local factor models of all countries outside the respective region. The regression equation is fitted with GMM. Confidence intervals are calculated using Newey and West (1987) adjusted standard errors using 6 lags. The number of observations for the regressions is fixed for the different factor models in each country. I require at least 120 observations for a test portfolio to be included. The "IDs" column holds the number of portfolios for which the asset pricing test is conducted. "R^2" hold the average goodness-of-fit measure of the regressions. "$\bar{\alpha}_c$ holds the average $\alpha_{c,i}$. "F_{GRS}" holds the Gibbons et al. (1989) test statistic for the regression. "Local", "Regional", and "Rest of World" hold the results for the respective factor models. Statistical significance at confidence levels of 10%, 5%, and 1% are indicated with *, **, and *** respectively.

			Local			*Regional*			*Rest of world*		
Country	Region	IDs	R^2	$\bar{\alpha}_c$	F_{GRS}	R^2	$\bar{\alpha}_c$	F_{GRS}	R^2	$\bar{\alpha}_c$	F_{GRS}
AUS	Established	10	78%	1.47%	0.58	60%	1.55%	0.58	51%	1.95%	0.57
AUT	Established	3	77%	1.15%	1.15	54%	5.56%	1.99	44%	2.60%	0.25
BEL	Established	7	70%	1.39%	0.59	57%	2.29%	0.43	41%	2.28%	0.57
CAN	Established	10	65%	2.35%	1.43	58%	2.22%	1.16	39%	4.72%*	1.53
CHE	Established	7	76%	2.08%	1.29	64%	1.69%	0.54	46%	5.49%	1.51
DEU	Established	9	73%	4.18%***	3.45***	64%	4.47%*	2.92***	44%	3.10%	1.55
DNK	Established	5	69%	2.48%	0.53	53%	2.74%	0.90	38%	6.08%*	2.34**
ESP	Established	6	70%	2.58%	1.43	52%	2.38%	0.76	38%	2.77%	1.99*
FIN	Established	5	68%	4.27%***	4.41***	58%	3.45%*	2.41**	42%	2.78%	0.53
FRA	Established	10	76%	2.47%	1.27	68%	3.49%	2.98***	45%	3.88%	1.83*
GBR	Established	10	74%	2.64%***	2.86***	63%	2.89%	1.33	44%	2.39%	1.44
ITA	Established	8	74%	2.11%	0.90	57%	2.72%	1.26	39%	3.58%	2.07**
NLD	Established	5	76%	3.23%*	2.05*	64%	3.40%**	2.45**	43%	4.75%**	2.39**
NOR	Established	8	72%	4.52%***	3.57***	59%	4.62%	2.51**	45%	4.68%	1.70*
SWE	Established	8	72%	1.80%	1.16	63%	1.87%	0.40	42%	4.64%	1.81*
USA	Established	10	76%	2.06%*	1.71*	62%	4.95%***	5.55***	41%	5.71%***	2.97***
Avg.		8	73%	2.55%	1.77	60%	3.14%	1.76	43%	3.84%	1.57

Table 3: *(continued)*

Country	Region	IDs	Local R^2	Local $\bar{\alpha}_c$	Local F_{GRS}	Regional R^2	Regional $\bar{\alpha}_c$	Regional F_{GRS}	Rest of world R^2	Rest of world $\bar{\alpha}_c$	Rest of world F_{GRS}
CHN	Emerged	9	86%	3.12%***	3.47***	22%	5.21%**	2.38**	12%	5.39%	1.57
HKG	Emerged	10	73%	3.81%***	1.90**	51%	4.06%*	1.99**	40%	2.46%	1.38
IDN	Emerged	8	73%	5.32%**	2.31**	26%	5.13%	1.56	31%	5.98%	1.64
JPN	Emerged	10	68%	3.15%***	2.74***	18%	2.37%	1.23	28%	1.61%	0.76
KOR	Emerged	10	73%	3.17%**	2.82***	40%	3.28%	0.99	45%	3.50%	1.76*
MYS	Emerged	10	79%	1.95%	1.18	33%	3.07%*	1.68*	37%	2.67%	1.33
NZL	Emerged	6	76%	3.36%**	3.88***	36%	3.30%*	3.31***	49%	2.81%	2.17**
PER	Emerged	3	82%	3.61%*	2.35*	32%	2.98%	0.75	30%	8.53%	1.18
PHL	Emerged	8	66%	1.74%	0.40	25%	3.15%	0.56	28%	2.48%	0.58
SGP	Emerged	8	79%	3.51%***	4.30***	48%	3.74%***	3.39***	54%	4.34%***	4.45***
THA	Emerged	8	77%	4.99%***	5.11***	33%	6.13%***	3.48***	38%	4.24%***	3.80***
TWN	Emerged	7	80%	1.25%	0.54	30%	2.80%	1.13	32%	2.18%	0.76
Avg.		8	76%	3.25%	2.58	33%	3.77%	1.87	35%	3.85%	1.78
ARG	Developing	3	81%	7.18%	2.24*	18%	6.33%	0.77	12%	2.11%	0.05
BRA	Developing	6	84%	2.67%	1.12	45%	5.09%	1.58	48%	2.19%	0.21
CHL	Developing	6	84%	3.16%***	4.59***	41%	3.23%***	3.54***	38%	2.56%	2.05*
GRC	Developing	6	82%	2.97%	1.12	31%	7.48%*	2.18**	37%	6.73%**	1.90*
IND	Developing	10	81%	3.19%*	1.66*	35%	7.55%	1.36	35%	7.24%	1.17
MEX	Developing	5	82%	2.69%	2.32**	44%	2.58%	0.68	53%	1.43%	0.15
TUR	Developing	6	83%	3.58%**	2.88***	30%	2.94%	0.52	29%	11.61%	0.85
Avg.		6	83%	3.63%	2.28	35%	5.03%	1.52	36%	4.84%	0.91
Avg.		**7**	**76%**	**3.01%**	**2.15**	**45%**	**3.73%**	**1.75**	**39%**	**4.04%**	**1.51**

over to a setting where portfolios are aggregated using equal-weighting instead of value-weighting in (3) and (4) for which results are stored in Table A.4 and Figure A.6[20], as well as for the traditional double sorting factor construction methodology following Fama and French (2017) for which results are reported in Table A.5 and Figure A.7[21] and the common sorting aggregation of regional factors (e.g., Fama & French, 2012; Hollstein, 2022) for which results are reported in Table A.6 and Figure A.8.[22] Lastly, the choice of factor model does

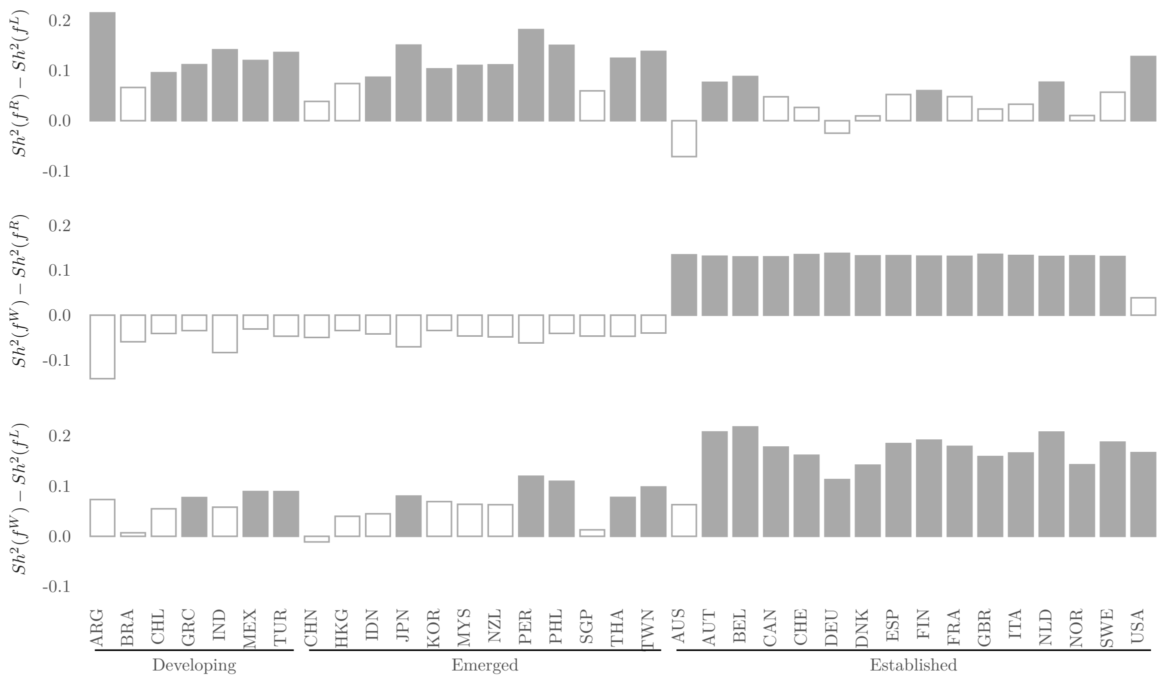

Figure 7: Barillas and Shanken (2017) test for Sharpe ratio comparison

The plot illustrates the results of the Barillas and Shanken (2017) squared Sharpe ratio test for local, regional, and rest-of-world Fama and French (2018) six-factor model returns in USD between 1992 and 2023. The local model consists of sorted long-short stock portfolios in the respective country. The regional model is a value-weighted average of local factor models in the respective region, excluding the respective country. The rest-of-world model is a value-weighted average of local factor models of all countries outside the respective region. The upper plot evaluates the regional model against the local model, i.e., $Sh^2(f^R) - Sh^2(f^L)$. The middle plot evaluates the rest-of-world model against the regional model, i.e., $Sh^2(f^W) - Sh^2(f^R)$. The lower plot evaluates the rest-of-world model against the local model, i.e., $Sh^2(f^W) - Sh^2(f^L)$. Bars are filled, if the difference between $Sh^2(f)$ is statistically significant at 5% confidence.

not alter the results either as shown by the regionally aggregated summaries in Table A.7.

the availability of data is highly heterogeneous among the countries in the sample.

[20]The overall picture does not change but there are two notable differences for the equally-weighted setting. First, regional portfolios perform worse than local portfolios concerning F_{GRS}. This is because regional models have much higher alphas than local models. Still, the most global rest-of-world model performs best with regard to this measure. Second, regional (rest-of-world) models have much higher statistically significant Sharpe ratios than the local model in the established (developing and emerged) region. This is caused by the large market capitalization of the U.S. that makes the respective models more concentrated in the value-weighted setting, reducing cross-country diversification among factors.

[21]This double-sorting methodology requires more data for the factor construction. Because of this, the sample size decreases to 21 countries for this methodology.

[22]The overall picture does not change but there is one notable differences for the sorting aggregation setting. The dominance of the regional and rest-of-world models over local models concerning higher statistically significant Sharpe ratios is decreased. This is presumably caused by the fact that the sorting aggregation does not ensure that all countries within a region are represented in the regional and rest-of-world portfolios, leaving less potential for diversification. However, the more global models still show higher Sharpe ratios than the local model in most countries.

4.3 Factor Importance

Given that employing different performance metrics leads to favoring different models, this analysis will examine the importance of various factors when optimizing asset pricing models for different performance metrics. For this purpose, I employ a modified version of Swade et al.'s (2024) iterative selection algorithm. The algorithm iteratively adds one new factor to an existing factor model and checks the performance of the augmented factor model. It always adds the factor that improves the model's performance for desired performance metrics the most. Once all possible factors have been tested and the best factor has been chosen, the factor is permanently added to the factor model, and the procedure is repeated until a certain number of factors in the model is reached or a convergence criterion is met. Formally, the algorithm works as follows:

1. Set $l \coloneqq 0$.
2. Test $K-l$ different augmented factor models that each add one of the remaining factors, labeled f_{test}, to the model from the previous iteration as

$$r_i = \alpha_i + \sum_{k=1}^{l} \beta_k f_k + \beta_{test} f_{test} + \varepsilon_i$$

to price returns from test assets $r_{i,t}$. Note that for the first iteration there are no factors in f_k.

3. Select the strongest model based on the lowest (highest) $\bar{\alpha}_c$, F_{GRS} (R^2) statistic.
4. Check whether improvement in the performance measure converged. If yes, set $l \coloneqq l+1$ and continue with step 2. If not, break.

As the convergence criterion, I assess whether the performance metric improved by 1% in comparison to the last iteration for R^2 and $\bar{\alpha}_c$. For the F_{GRS} I set the required performance improvement to 5% from the previous iteration.

The results for the analysis are illustrated in Figure 8. The market factor is highlighted because it is the only one that is not a long-short factor and is thus substantially different than the others. In line with previous results, the selection of the optimal factor model is vastly different for different metrics. With regards to R^2, we see that by far the most important factor is the local market factor. Furthermore, to improve R^2 the most, almost exclusively local factors are selected. For $\bar{\alpha}_c$ the result is quite different. It can be observed that a market factor, either local or regional is most important for reducing alphas. In terms of non-market factors, roughly the same number of factors are selected to reduce alphas (13 local, 15 regional, and 15 rest-of-world). Also, we can see a strong divergence between countries regarding what factors are selected. For F_{GRS} the result looks similar to the one for $\bar{\alpha}_c$. Again, some market factor is selected among the first factors.

As a last test to determine what factors are important, I follow the procedure by Barillas and Shanken (2018). They argue that the best factor asset pricing model among a pool of potential factors must be able to price all factors that are not part of the model. They propose a Bayesian procedure that simultaneously assesses what combination of factors is most likely to price all others. Their testing framework has been refined by Chib et al. (2020) who show that the original version uses an improper prior.

In short, for a set of O traded potential risk factors, there are $J = 2^O - 1$ possible factor combinations. The model space of possible factor models is $\mathcal{M} = \{\mathbb{M}_1, \mathbb{M}_2, \ldots, \mathbb{M}_J\}$. $\mathbb{M}_j$ is a model of P_j included factors $\tilde{f}_j$ and $O - P_j$ excluded factors f_j^*. The data-generating process for the two sets of factors is defined as

$$\tilde{f}_j = \tilde{\alpha}_j + \tilde{\epsilon}_j \tag{10}$$

$$f_{j,t}^* = B_{j,f}^* \tilde{f}_j + \epsilon_j^* \tag{11}$$

where $\tilde{\epsilon}_j$ and ϵ_j^* are multivariate normally distributed residual vectors. The log marginal likelihood of $\mathbb{M}_j (j \neq J)$ with sample data y can be calculated in closed form as

$$\log \tilde{m}(y|\mathbb{M}_j) = \log \tilde{m}(\tilde{f}|\mathbb{M}_j) + \log \tilde{m}(f^*|\mathbb{M}_j) \tag{12}$$

for which the detailed solution is described in Appendix D. I calculate $\log \tilde{m}(y|\mathbb{M}_j)$ for all $\mathbb{M}_j$. The model with the highest log marginal likelihood is the winning model as all other terms can be summarized in a normalization constant.

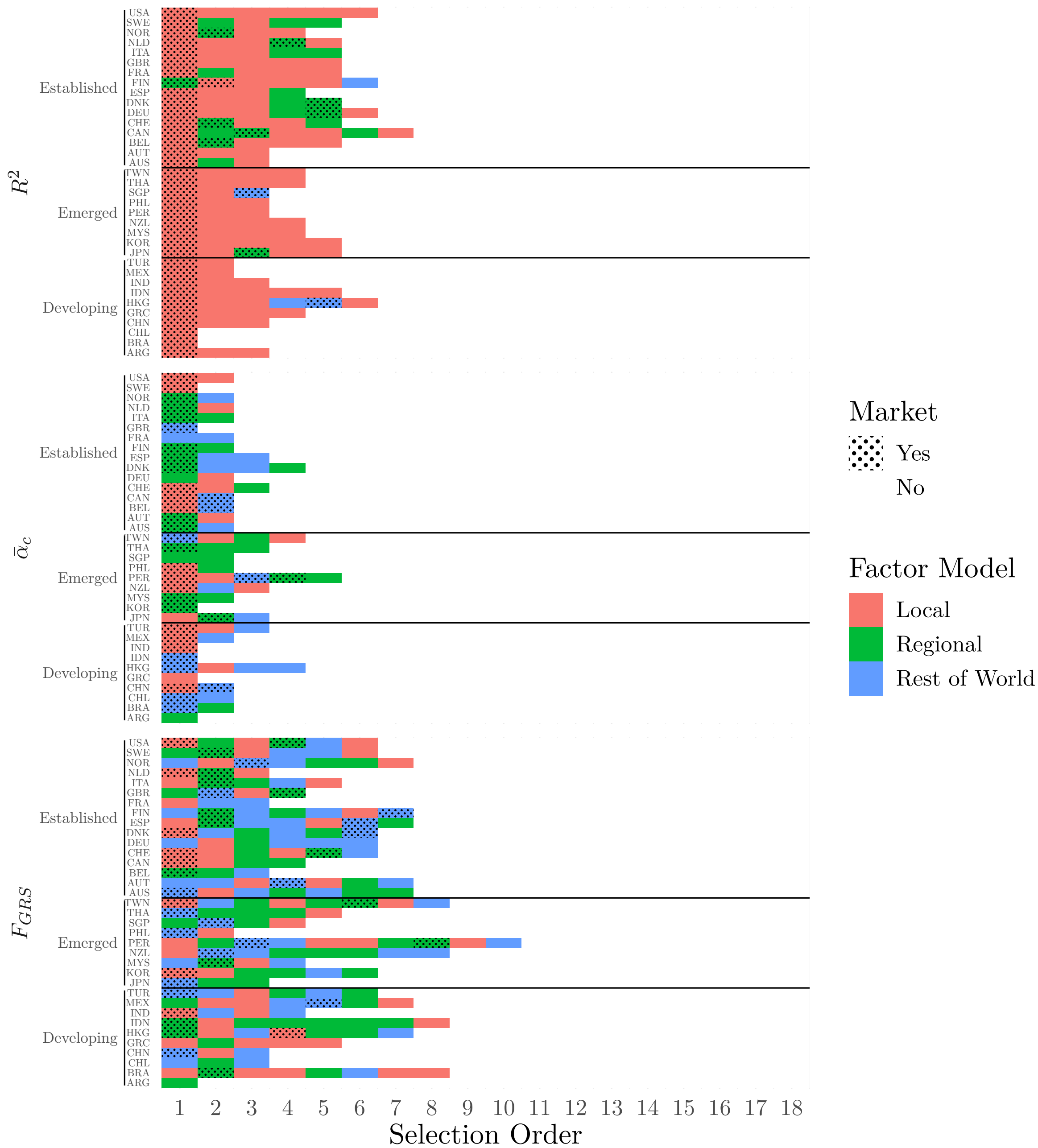

Figure 8: Factor selection following Swade et al. (2024) for local, regional, and rest-of-world factors

The plot illustrates the factor selection using the Swade et al. (2024) algorithm on local, regional, and rest-of-world Fama and French (2018) six-factor model returns in USD between 1992 and 2023. The algorithm augments a factor model $\sum_{k=1}^{l} f_k$ by adding each factor from the test pool and evaluates the model's performance in pricing GICS 10 industry portfolio returns, r_i as test assets as $r_i = \alpha_i + \sum_{k=1}^{l} \beta_k f_k + \beta_{test} f_{test} + \varepsilon_i$. I require at least 120 observations for a test portfolio to be included in the test. The factor f_{test} that improves performance the most is added to $\sum_{k=1}^{l} f_k$, and $l = l + 1$. Initially, $l \coloneqq 0$ and $\sum_{k=1}^{l} f_k$ is empty. The process continues until performance converges. The plots show results for models optimized to increase R^2 (top), decrease average α_i, $\bar{\alpha}_c$ (middle), and decrease the Gibbons et al. (1989) statistic, F_{GRS} (bottom). Convergence is achieved if the performance measure does not improve by 1% for R^2 and $\bar{\alpha}_c$, or by 5% for F_{GRS} compared to the last iteration.

Following Chib et al. (2020) I run the scanning process for all J models using 10% of the available data as the training sample. The results are plotted in Figure 9 where the model with the highest $\log \tilde{m}(y|\mathbb{M}_j)$ is illustrated for each country. Since the scanning algorithm can pick different numbers of factors for each country, the factor contribution is shown on a relative scale. The total number of selected factors is illustrated on the right side of the graph. Overall, the results show that the algorithm picks more rest-of-world factors for the established region and more regional factors for the two other regions. While a few local factors are chosen for some countries, all models consist of more global than local factors.

Figure 9: Bayesian factor selection for local, regional, and rest-of-world factors

The plot illustrates the factor selection using the Bayesian scanning by Chib et al. (2020) which is based on the works of (Barillas & Shanken, 2018). The data used for the scan are local, regional, and rest-of-world Fama and French (2018) six-factor model returns in USD between 1992 and 2023. The algorithm tests what set of factors among a pool of potential risk factors is most likely to price all remaining factors. The plot illustrates the composition of the most likely model for every country. It is highlighted whether a factor is local, regional, or rest-of-world and whether it is a market factor. The total number of factors chosen is documented on the left side of the plot.

4.4 Controlling for Local Factors

The previous selection analysis shows that the local market factor is particularly important for improving R^2 and also plays a vital role for reducing $\bar{\alpha}_c$ in many countries. Given the importance of these two performance measures and the unique nature of the local market factor as the only factor with directional exposure, it is essential to analyze whether the findings of previous studies—which suggest that local factor models perform best (Fama & French, 2012; Griffin, 2002; Hollstein, 2022)—are driven by the local market factor. This analysis is detailed in the following paragraph.

I repeat the asset pricing analysis from Section 4.2 using the same regression equation as in (7) but now all three-factor models are equipped with the local factor model. In other words, I swap the regional and rest-of-world market factor for the local market factor in the regional and rest-of-world model respectively. All other specifications of the testing procedure remain as before.

A summary of the results is documented in Table 4 that compares the results to the setting where each model uses its own market factor. Complete country-level results are reported in Table A.8 and results for the Sharpe ratio test are illustrated in Figure A.9. We can see that the performance for $\bar{\alpha}_c$ and R^2 of the regional and rest-of-world global models improves strongly when granted the local market factor. Still, the local model does better than the more global models with regards to these two metrics but the difference is far smaller than for the case when all models use their own market factor. With regards to the F_{GRS} statistic, the more global models still perform better than the local model even though their F_{GRS} slightly increases. Similarly, the evaluation based on $Sh^2(f)$ remains unchanged as more global models outperform their local counterparts. These results also carry over to other factor models for which the results are stored in Table A.7.

Lastly, I control whether the pricing performance of the regional and rest-of-world factors can be explained by exposure to their regional or local counterparts. In a similar fashion to the structural analysis, where I regress all factors on the market factor and performed the analysis using the residuals, I now do the same but using three different settings:

Table 4: Asset pricing test comparison for local vs. regional vs. rest-of-world factor model with GICS portfolios using local market factor in every model

The table holds the regionally aggregated results of asset pricing regressions for local, regional, and rest-of-world factor model returns between 1992 and 2023. The regression setup is $r_{c,i} = \alpha_{c,i} + \sum \beta_{c,i} f_c^{L/R/W} + \varepsilon_{c,i}$, for each country, c. $r_{c,i}$ are return of industry portfolio i for GICS 10 portfolios. $f_c^{L/R/W}$ contains the returns of a local, regional or rest-of-world factor model. The local model consists of sorted long-short stock portfolios in the respective country. The regional model is a value-weighted average of local factor models in the respective region, excluding the respective country. The rest-of-world model is a value-weighted average of local factor models of all countries outside the respective region. The models are evaluated once using their market factor (columns 3-6) and once where the market factor is fixed to be the market factor of the local model for the regional and rest-of-world models (columns 7-10). The regression equation is fitted using GMM. Confidence intervals are calculated using Newey and West (1987) adjusted standard errors with 6 lags. The number of observations for the regressions is fixed for the different factor models in each country. I require at least 120 observations for a test portfolio to be included. The "Region" column shows over what region the results are aggregated. The "Models" column shows whether the aggregated results are for a local, regional, or rest-of-world model. "R^2" hold the average goodness-of-fit measure of the regressions. "$\bar{\alpha}_c$ holds the average $\alpha_{c,i}$. "F_{GRS}" holds the Gibbons et al. (1989) test statistic for the regression. "ΔSh^2" holds the results of the Barillas and Shanken (2017) squared Sharpe ratio test. Results in this column are always evaluated against the local model. E.g., for the regional model, the column holds the result from $Sh^2(f^R) - Sh^2(f^L)$. "Local", "Regional", and "Rest of World" hold the results for the respective factor models.

		Model Market Factor				*Local Market Factor*			
Region	Model	R^2	$\bar{\alpha}_c$	F_{GRS}	ΔSh^2	R^2	$\bar{\alpha}_c$	F_{GRS}	ΔSh^2
Developing	Local	83%	3.63%	2.28		83%	3.63%	2.28	
	Regional	35%	5.03%	1.52	0.15	78%	4.59%	2.18	0.11
	Rest of World	36%	4.84%	0.91	0.06	78%	4.33%	1.80	0.04
Established	Local	73%	2.55%	1.77		73%	2.55%	1.77	
	Regional	60%	3.14%	1.76	0.10	70%	2.90%	1.80	0.11
	Rest of World	43%	3.84%	1.57	0.07	68%	2.95%	1.60	0.01
Emerged	Local	76%	3.25%	2.58		76%	3.25%	2.58	
	Regional	33%	3.77%	1.87	0.04	71%	3.47%	2.13	0.02
	Rest of World	35%	3.85%	1.78	0.17	72%	2.94%	2.15	0.16
Avg.	**Local**	**76%**	**3.01%**	**2.15**		**76%**	**3.01%**	**2.15**	
	Regional	**45%**	**3.73%**	**1.75**	**0.08**	**72%**	**3.44%**	**1.99**	**0.07**
	Rest of World	**39%**	**4.04%**	**1.51**	**0.11**	**71%**	**3.22%**	**1.83**	**0.09**

$$f_{c,\tau}^{R} = \alpha_{c,\tau} + \beta_{c,\tau} f_{c,\tau}^{L} + \varepsilon_{c,\tau,R|L} \tag{13}$$

$$f_{c,\tau}^{W} = \alpha_{c,\tau} + \beta_{c,\tau} f_{c,\tau}^{L} + \varepsilon_{c,\tau,W|L} \tag{14}$$

$$f_{c,\tau}^{W} = \alpha_{c,\tau} + \beta_{c,\tau} f_{c,\tau}^{R} + \varepsilon_{c,\tau,W|R} \tag{15}$$

In (13) I clean the regional factors for exposure to the respective local factors. In (14) I clean the rest-of-world factors for exposure to the respective local factors. Lastly, in (15) I remove exposure to regional factors from the respective rest-of-world factors. Using the residuals from the above regressions I perform the asset pricing analysis. The results of the asset pricing analysis remain almost unchanged and are therefore not reported. Additionally, I compute the correlations between the cleaned residuals and the respective factors before cleaning in Table 5. We can see that the correlations between the clean residuals and their respective factors are very high, indicating that little of their factor exposure can be explained by exposure to less global factors.

Table 5: Residual correlations with respective factors

The table holds correlations between factors and regression residuals of local, regional, and rest-of-world Fama and French (2018) six-factor model returns in USD between 1992 and 2023. The regression following three regressions are executed for each factor, τ: $f^R_{c,\tau} = \alpha_{c,\tau} + \beta_{c,\tau} f^L_{c,\tau} + \varepsilon_{c,\tau,R|L}$ cleans regional factor returns $f^R_{c,\tau}$ for local factor return exposure. The Pearson correlation between $f^R_{c,\tau}$ and $\varepsilon_{c,\tau,R|L}$ is reported in the second column. $f^W_{c,\tau} = \alpha_{c,\tau} + \beta_{c,\tau} f^L_{c,\tau} + \varepsilon_{c,\tau,W|L}$ cleans rest-of-world factor returns $f^W_{c,\tau}$ for local factor return exposure. The Pearson correlation between $f^W_{c,\tau}$ and $\varepsilon_{c,\tau,W|L}$ is reported in the third column. $f^W_{c,\tau} = \alpha_{c,\tau} + \beta_{c,\tau} f^R_{c,\tau} + \varepsilon_{c,\tau,W|R}$ cleans rest-of-world factor returns $f^W_{c,\tau}$ for regional factor return exposure. The Pearson correlation between $f^W_{c,\tau}$ and $\varepsilon_{c,\tau,W|R}$ is reported in the fourth column.

Factor	$cor(f^R_c, \varepsilon_{c,R\|L})$	$cor(f^W_c, \varepsilon_{c,W\|L})$	$cor(f^W_c, \varepsilon_{c,W\|R})$
Market	67%	74%	64%
Size	96%	99%	99%
Value	87%	92%	80%
Investment	93%	97%	91%
Profitability	98%	99%	100%
Momentum	85%	91%	79%

5 Conclusion

In this study, I revisit whether assets are priced on a local, regional, or global level. Previous research has largely converged on the superiority of local asset pricing models over their regional and global counterparts (Fama & French, 2012, 2017; Griffin, 2002; Hollstein, 2022). However, these studies often make simplifying assumptions, such as uniform regionality across factors, arbitrary regional definitions, limited evaluation metrics, and isolated model testing, neglecting the nuanced interplay of local, regional, and global components in asset pricing.

In this paper, I address these limitations and offer fresh insights into the mechanics of factor asset pricing. Leveraging principal component analysis, I uncover varying degrees of regional integration, with clustering patterns consistent across factors but exhibiting diverse strengths. Through a data-driven approach, I define three distinct clusters, primarily delineated along continental and developmental lines, that lay the groundwork for constructing regional asset pricing models. Also, I do not find evidence to consider Japan as an outlier warranting the creation of its own distinct region.

Drawing from these structural insights, I conduct rigorous asset pricing tests using diverse metrics. While local models perform best in terms of R^2 and alpha reduction, more global models outperform when considering Sharpe ratios. Notably, the dominance of local models in certain performance measures is largely attributable to the influence of the local market factor. When creating optimal asset pricing models from a pool of local, regional, and rest-of-world factors, I show that the winning models consistently include elements of local, regional, and rest-of-world factors.

Overall, my findings underscore the complexity of the asset pricing landscape, challenging simplistic characterizations of local versus global dynamics. The answer to the asset pricing conundrum is contingent upon the choice of performance metrics, with an ideal model likely comprising a blend of local, regional, and global components. These insights bridge empirical asset pricing literature which mostly advocates local models with recent findings in the prediction literature, that suggests a convergence towards more globally oriented models in predicting asset returns.

References

Abdi, H., Williams, L. J., & Valentin, D. (2013). Multiple factor analysis : principal component analysis for multitable and multiblock data sets. *WIREs Computational Statistics*, *5*(2), 149–179.

Arce, M., & Mora, A. (2002). Empirical evidence of the effect of European accounting differences on the stock market valuation of earnings and book value. *European Accounting Review*, *11*(3), 573–599.

Asness, C. S., Ilmanen, A., Israel, R., & Moskowitz, T. J. (2015). Investing With Style. *Journal of Investment Management*, *13*(1), 27–63.

Ball, R., Gerakos, J., Linnainmaa, J. T., & Nikolaev, V. (2016). Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics*, *121*(1), 28–45.

Banerjee, A., & Davé, R. N. (2004). Validating clusters using the Hopkins statistic. *IEEE International Conference on Fuzzy Systems*, *1*, 149–153.

Barillas, F., Kan, R., Robotti, C., & Shanken, J. (2020). Model Comparison with Sharpe Ratios. *Journal of Financial and Quantitative Analysis*, *55*(6), 1840–1874.

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

Barillas, F., & Shanken, J. (2018). Comparing Asset Pricing Models. *Journal of Finance*, *73*(2), 715–754.

Bekaert, G., Harvey, C. R., Kiguel, A., & Wang, X. (2016). Globalization and Asset Returns. *Annual Review of Financial Economics*, *8*, 221–288.

Brandt, M. W., Cochrane, J. H., & Santa-Clara, P. (2006). International risk sharing is better than you think, or exchange rates are too smooth. *Journal of Monetary Economics*, *53*(4), 671–698.

Brooks, R., & Del Negro, M. (2005). Country versus region effects in international stock returns. *Journal of Portfolio Management*, *31*(4), 67–72.

Chib, S., Zeng, X., & Zhao, L. (2020). On Comparing Asset Pricing Models. *The Journal of Finance*, *75*(1), 551–577.

Choi, D., Jiang, W., & Zhang, C. (2022). Alpha Go Everywhere: Machine Learning and International Stock Returns. *SSRN Electronic Journal*.

Detzel, A., Novy-Marx, R., & Velikov, M. (2023). Model Comparison with Transaction Costs. *Journal of Finance*, *78*(3), 1227–1831.

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

Escofier, B., & Pagès, J. (1994). Multiple factor analysis. *Computational Statistics & Data Analysis*, *18*(1), 121–140.

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

Fama, E. F., & French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics*, *105*(3), 457–472.

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

Fama, E. F., & French, K. R. (2017). International tests of a five-factor asset pricing model. *Journal of Financial Economics*, *123*(3), 441–463.

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

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A Test of the Efficiency of a Given Portfolio. *Econometrica*, *57*(5), 1121–1152.

Grauer, F. L., Litzenberger, R. H., & Stehle, R. E. (1976). Sharing rules and equilibrium in an international capital market under uncertainty. *Journal of Financial Economics*, *3*(3), 233–256.

Griffin, J. M. (2002). Are the Fama and French Factors Global or Country Specific? *Review of Financial Studies*, *15*(3), 783–803.

Gupta, T., & Kelly, B. (2019). Factor Momentum Everywhere. *The Journal of Portfolio Management*, *45*(3), 13–36.

Hau, H. (2011). Global versus local asset pricing: A new test of market integration. *Review of Financial Studies*, *24*(12), 3891–3940.

Hellum, O., Pedersen, L. H., & Rønn-Nielsen, A. (2023). How Global is Predictability? The Power of Financial Transfer Learning. *SSRN Electronic Journal*.

Hollstein, F. (2022). Local, Regional, or Global Asset Pricing? *Journal of Financial and Quantitative Analysis*, *57*(1), 291–320.

Hopkins, B. (1954). A new method for determining the type of distribution of plant individuals. *Annals of Botany*, *18*(2), 213–227.

Horn, J. L. (1965). A Rationale and Test for the Number of Factors in Factor Analysis. *Psychometrika*, *30*(2), 179–185.

Hou, K., Karolyi, G. A., Kho, B.-c., Hou, K., & Karolyi, G. A. (2011). What Factors Drive Global Stock Returns? *Review of Financial Studies*, *24*(8), 2527–2574.

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

Ilmanen, A. (1995). Time-Varying Expected Returns in International Bond Markets. *The Journal of Finance*, *50*(2), 481–506.

Jensen, T. I., Kelly, B. T., & Pedersen, L. H. (2023). Is There a Replication Crisis in Finance? *The Journal of Finance*, *78*(5), 2465–2518.

Karolyi, G. A., & Stulz, R. M. (2003). Are financial assets priced locally or globally? *Handbook of the economics of finance* (1st ed., pp. 975–1020). Elsevier B.V.

Karolyi, G. A., & Wu, Y. (2018). A New Partial-Segmentation Approach to Modeling International Stock Returns. *Journal of Financial and Quantitative Analysis*, *53*(2), 507–546.

Kaufman, L., & Rousseeuw, P. J. (1990). Partitioning Around Medoids (Program PAM). *Finding groups in data: An introduction to cluster analysis* (pp. 68–125). Wiley & Sons Ltd.

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

Kelly, B. T., & Xiu, D. (2023). Financial Machine Learning. *University of Chicago Working Paper Series*, *2023-100*.

Lewis, K. K. (2011). Global asset pricing. *Annual Review of Financial Economics*, *3*, 435–466.

Longstaff, F. A., Pan, J., Pedersen, L. H., & Singleton, K. J. (2011). How sovereign is sovereign credit risk? *American Economic Journal: Macroeconomics*, *3*(2), 75–103.

Murtagh, F., & Legendre, P. (2014). Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*, *31*, 274–295.

Newey, W. K., & West, K. D. (1987). A Simple , Positive Semi-Definite , Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, *55*(3), 703–708.

Petzev, I., Schrimpf, A., & Wagner, A. F. (2016). Has the Pricing of Stocks Become More Global? *BIS Working Paper Series*, (560).

Rapach, D. E., Strauss, J. K., & Zhou, G. (2013). International Stock Return Predictability: What Is the Role of the United States? *Journal of Finance*, *68*(4), 1633–1662.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.

Simoens, M., & Vennet, R. V. (2021). Bank performance in Europe and the US: A divergence in market-to-book ratios. *Finance Research Letters*, *40*(November 2019), 101672.

Solnik, B. H. (1974). An equilibrium model of the international capital market. *Journal of Economic Theory*, *8*(4), 500–524.

Stulz, R. M. (1981). A model of international asset pricing. *Journal of Financial Economics*, *9*(4), 383–406.

Swade, A., Hanauer, M. X., Lohre, H., & Blitz, D. (2024). Factor Zoo (. zip). *The Journal of Portfolio Management*, *50*(3), 11–31.

Tang, T. M., & Allen, G. I. (2021). Integrated principal components analysis. *Journal of Machine Learning Research*, *22*, 1–71.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *63*(2), 411–423.

Tobek, O., & Hronec, M. (2021). Does it pay to follow anomalies research? Machine learning approach with international evidence. *Journal of Financial Markets*, *56*, 100588.

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244.

Appendices

A Additional Tables and Figures

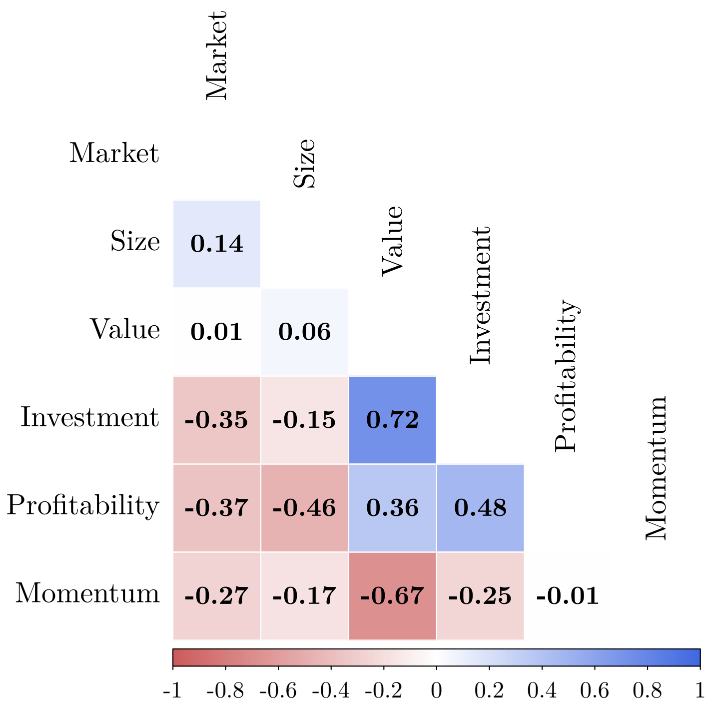

Figure A.1: Factor correlations

The figure shows the average Pearson correlation coefficient between factors from different themes or between factors from a theme and market returns on the off-diagonal elements. The diagonal elements show the average Pearson correlation coefficients for factors in the same theme. The underlying data are monthly world factors and market returns between January 1992 and December 2023. The world factor and market returns are calculated as capitalization-weighted averages of the country factor returns.

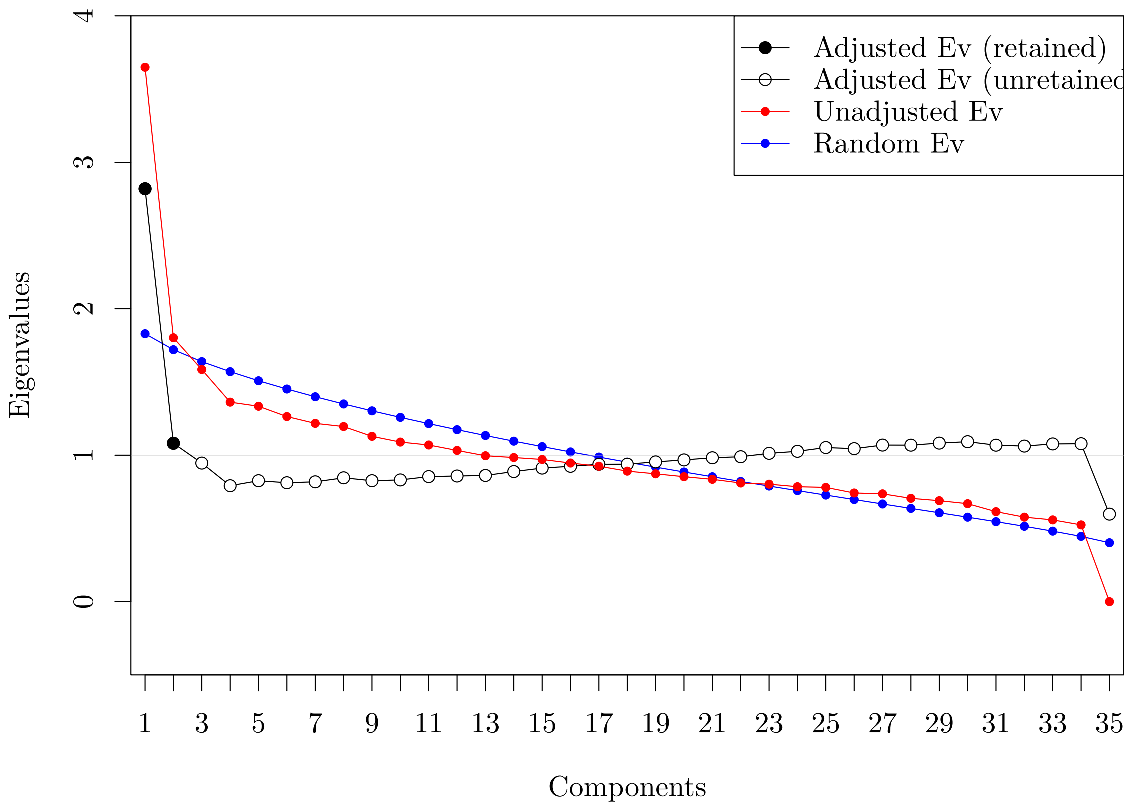

Figure A.2: Horn (1965) adjustment of eigenvalues

The figure results of the eigenvalue correction for eigenvalues that were created from international factor and market return correlation matrices using the iPCA algorithm by Tang and Allen (2021). The red line illustrates the unadjusted eigenvalues. The blue line illustrates the average eigenvalues from randomly generated correlation matrices of randomly generated data using 10.000 iterations. The black line illustrates the difference between the unadjusted eigenvalues and the randomly generated eigenvalues. Following the Kaiser (1960) criterion only adjusted eigenvalues that are greater than 1 are deemed to carry information about the correlation structure of the data.

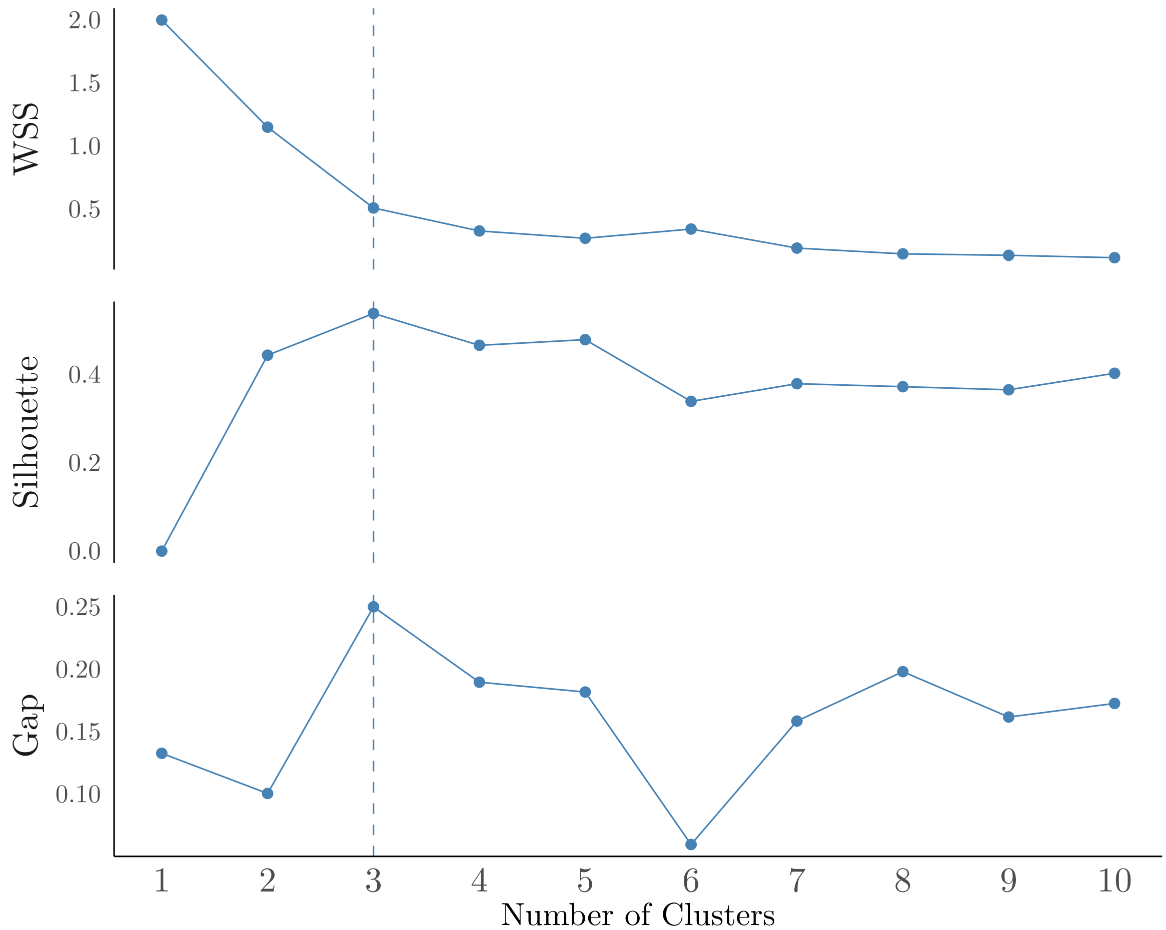

Figure A.3: Clustering statistics using the k-means algorithm

The figure holds three clustering measures for the first two eigenvectors, of correlation matrices from international Fama and French (2018) six-factor model returns in USD between 1992 and 2023 that were cleaned for market exposure. The eigenvectors are calculated jointly using the iPCA procedure by Tang and Allen (2021). ”WSS” is the average distance between points within a cluster. ”Silhouette” is the silhouette statistic by Rousseeuw (1987) which compares the distance of each data point within the cluster to the average distance to all other points outside the cluster. ”Gap” is the gap statistic by Tibshirani et al. (2001) which compares the average intra-cluster variation to the intra-cluster variation of uniformly randomly generated data. The x-axis holds the number of respective clusters for which the measures were calculated. The clusters are generated using k-means clustering.

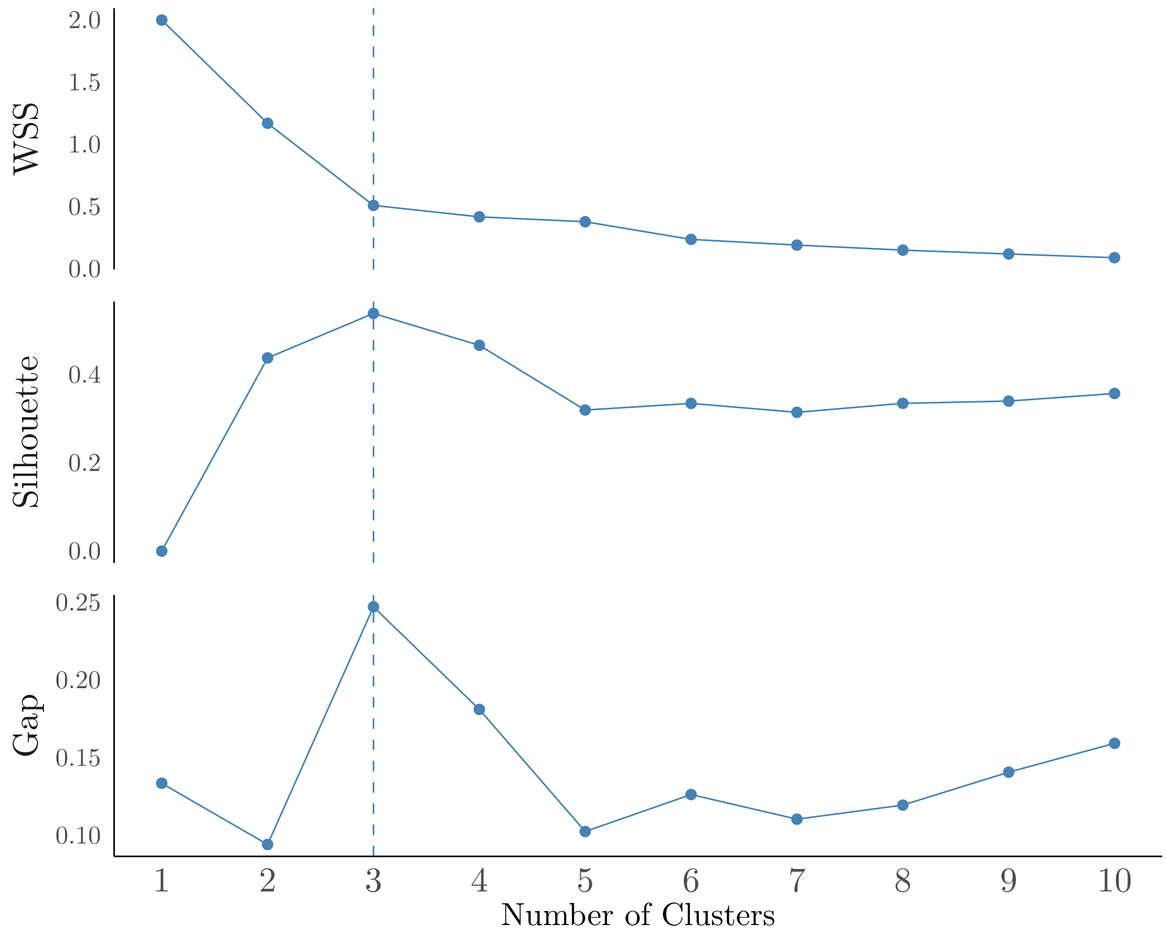

Figure A.4: Cluster statistics using the PAM algorithm

The figure holds three clustering measures for the first two eigenvectors, of correlation matrices from international Fama and French (2018) six-factor model returns in USD between 1992 and 2023. The eigenvectors are calculated jointly using the iPCA procedure by Tang and Allen (2021). ”WSS” is the average distance between points within a cluster. ”Silhouette” is the silhouette statistic by Rousseeuw (1987) which compares the distance of each data point within the cluster to the average distance to all other points outside the cluster. ”Gap” is the gap statistic by Tibshirani et al. (2001) which compares the average intra-cluster variation to the intra-cluster variation of uniformly randomly generated data. The x-axis holds the number of respective clusters for which the measures were calculated. The clusters are generated using the PAM clustering algorithm by Kaufman and Rousseeuw (1990).

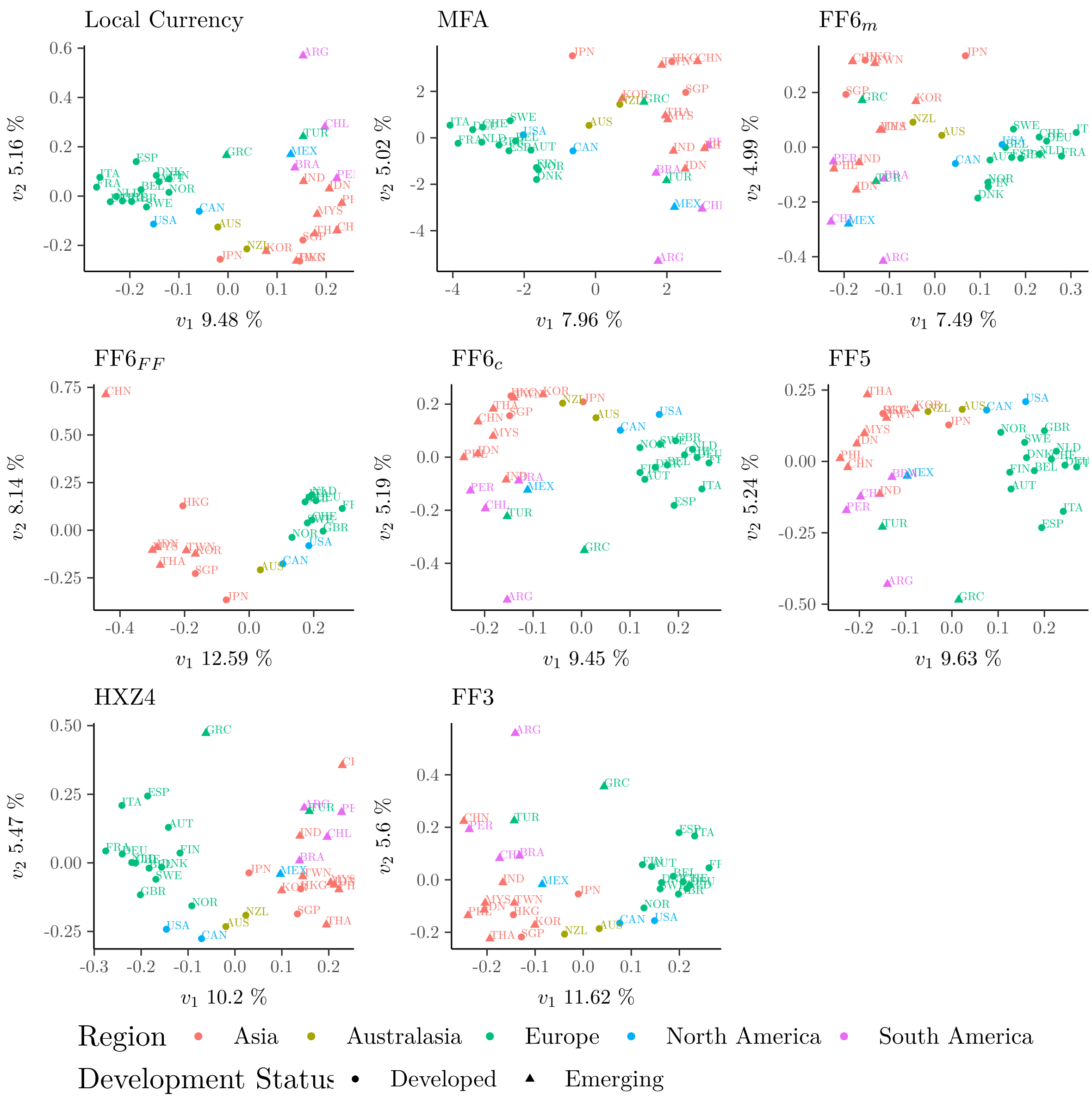

Figure A.5: Robustness of integrated principal component analysis of global factor returns
The plot illustrates the first two eigenvectors, v_1 and v_2, and the percentage of explained variance of their respective eigenvalues for international factor returns of the Fama and French (2018) six-factor model returns in USD that were decomposed using the iPCA algorithm by Tang and Allen (2021). Each plot illustrates a modified version of this baseline setting where one of the methodological decisions was changed. "Local Currency" shows the results if returns are in domestic currency instead of the USD. "MFA" displays the results of when the MFA algorithm was employed instead of the iPCA algorithm. The remaining plots display the results for different factor models. "FF6_m" displays the results for the Fama and French (2018) six-factor model from which the market factor was removed. "FF6_{FF}" displays results for the Fama and French (2018) six-factor model that was created using the traditional factor formation methodology by Fama and French (2017). "FF6_c" displays the results for the Fama and French (2018) six-factor model where the profitability factor was calculated using cash profitability. "FF5" shows the results for the Fama and French (2015) five-factor model, "HXZ4" displays the results for the Hou et al. (2015) q-factor model, and "FF3" displays the results for the Fama and French (1993) three-factor model.

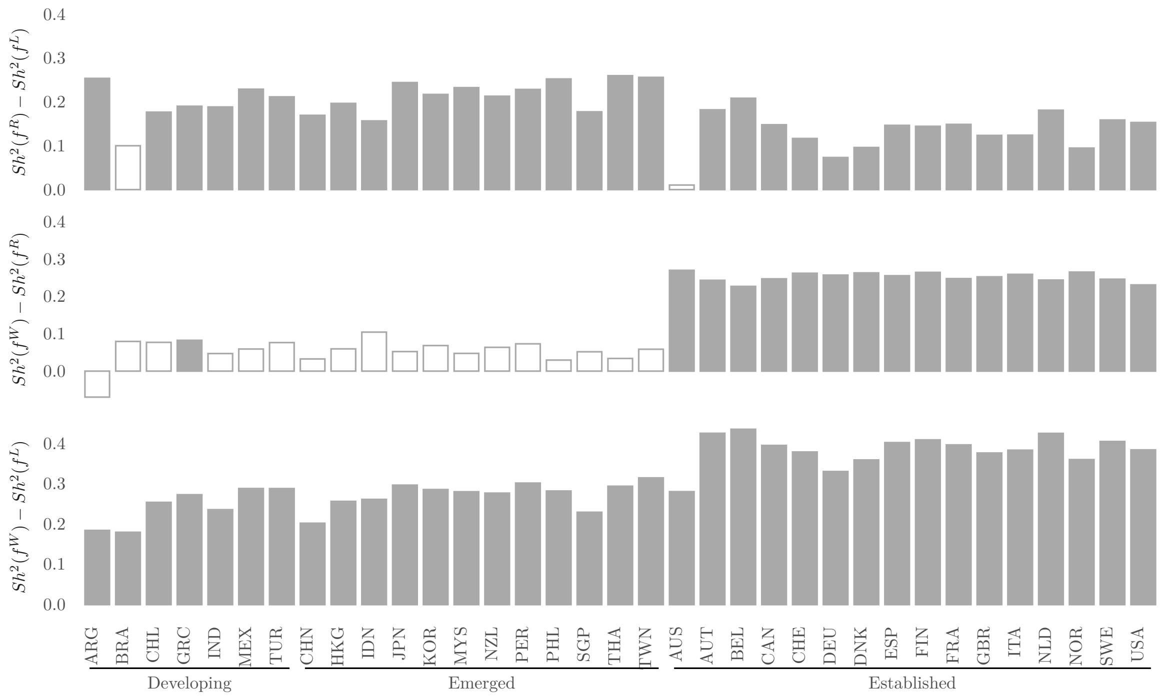

Figure A.6: Barillas and Shanken (2017) test for Sharpe ratio comparison with equally weighted regional and rest-of-world factors

The plot illustrates the results of the Barillas and Shanken (2017) squared Sharpe ratio test for local, regional, and rest-of-world Fama and French (2018) six-factor model returns in USD between 1992 and 2023. The local model consists of sorted long-short stock portfolios in the respective country. The regional model is an equally weighted average of local factor models in the respective region, excluding the respective country. The rest-of-world model is an equally weighted average of local factor models of all countries outside the respective region. The upper plot evaluates the regional model against the local model, i.e., $Sh^2(f^R) - Sh^2(f^L)$. The middle plot evaluates the rest-of-world model against the regional model, i.e., $Sh^2(f^W) - Sh^2(f^R)$. The lower plot evaluates the rest-of-world model against the local model, i.e., $Sh^2(f^W) - Sh^2(f^L)$. Bars are filled, if the difference between $Sh^2(f)$ is statistically significant at 5% confidence.

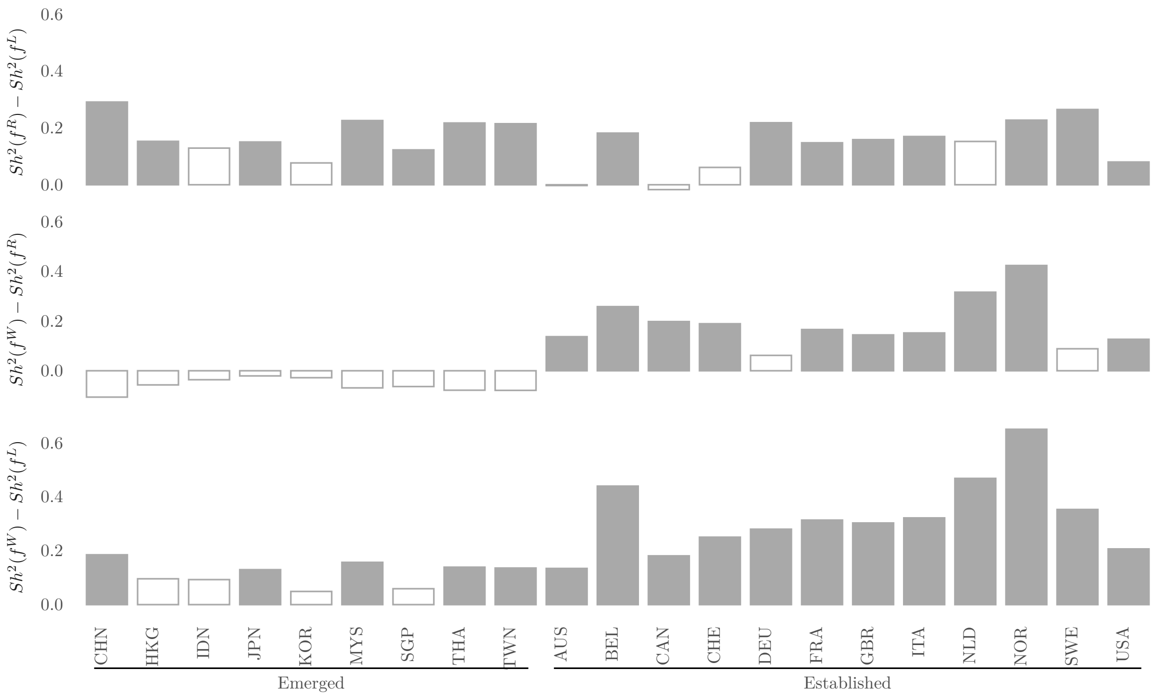

Figure A.7: Barillas and Shanken (2017) test for Sharpe ratio comparison with traditional Fama and French (2017) factor formation methodology

The plot illustrates the results of the Barillas and Shanken (2017) squared Sharpe ratio test for local, regional, and rest-of-world Fama and French (2018) six-factor model returns in USD between 1992 and 2023. The local model is built following the methodology by Fama and French (2017) that creates 2×3 sorted long-short stock portfolios in the respective country. The regional model is built from sorted long-short stock portfolios in the respective region, excluding stocks from the respective country. The rest-of-world model consists of sorted long-short stock portfolios with stocks from all countries outside the respective region. The upper plot evaluates the regional model against the local model, i.e., $Sh^2(f^R) - Sh^2(f^L)$. The middle plot evaluates the rest-of-world model against the regional model, i.e., $Sh^2(f^W) - Sh^2(f^R)$. The lower plot evaluates the rest-of-world model against the local model, i.e., $Sh^2(f^W) - Sh^2(f^L)$. Bars are filled, if the difference between $Sh^2(f)$ is statistically significant at 5% confidence.

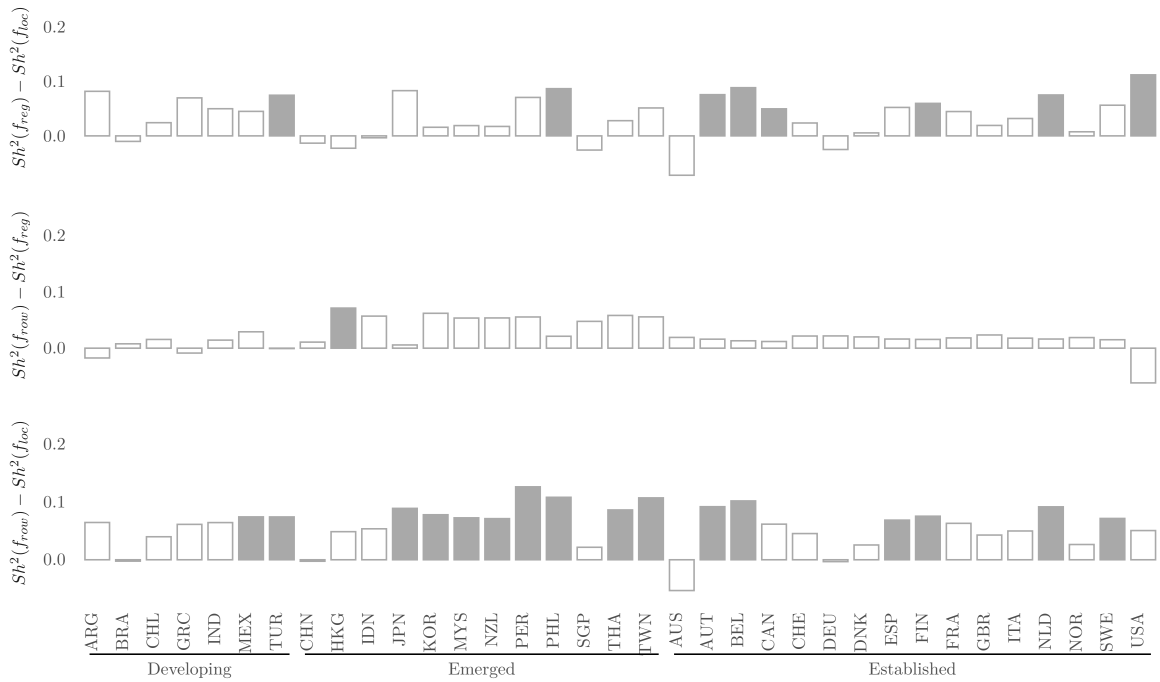

Figure A.8: Barillas and Shanken (2017) test for Sharpe ratio comparison with sorting aggregation procedure

The plot illustrates the results of the Barillas and Shanken (2017) squared Sharpe ratio test for local, regional, and rest-of-world Fama and French (2018) six-factor model returns in USD between 1992 and 2023. The local model consists of sorted long-short stock portfolios in the respective country. The regional model is built from sorted long-short stock portfolios in the respective region, excluding stocks from the respective country. The rest-of-world model consists of sorted long-short stock portfolios with stocks from all countries outside the respective region. The upper plot evaluates the regional model against the local model, i.e., $Sh^2(f^R) - Sh^2(f^L)$. The middle plot evaluates the rest-of-world model against the regional model, i.e., $Sh^2(f^W) - Sh^2(f^R)$. The lower plot evaluates the rest-of-world model against the local model, i.e., $Sh^2(f^W) - Sh^2(f^L)$. Bars are filled, if the difference between $Sh^2(f)$ is statistically significant at 5% confidence.

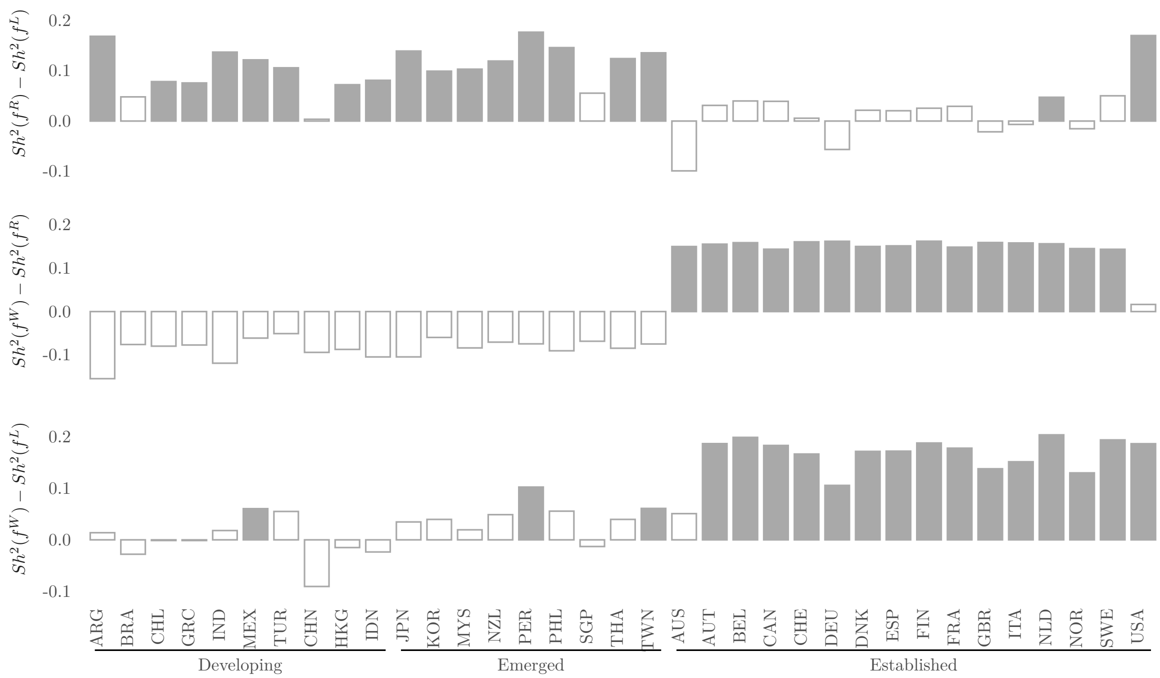

Figure A.9: Barillas and Shanken (2017) test for Sharpe ratio comparison with local market factor in every model

The plot illustrates the results of the Barillas and Shanken (2017) squared Sharpe ratio test for local, regional, and rest-of-world Fama and French (2018) six-factor model returns in USD between 1992 and 2023. The local model consists of sorted long-short stock portfolios in the respective country. The regional model is a value-weighted average of local factor models in the respective region, excluding the respective country. The rest-of-world model is a value-weighted average of local factor models of all countries outside the respective region. The market factor is fixed to be the market factor of the local model for all other models. The upper plot evaluates the regional model against the local model, i.e., $Sh^2(f^R) - Sh^2(f^L)$. The middle plot evaluates the rest-of-world model against the regional model, i.e., $Sh^2(f^W) - Sh^2(f^R)$. The lower plot evaluates the rest-of-world model against the local model, i.e., $Sh^2(f^W) - Sh^2(f^L)$. Bars are filled, if the difference between $Sh^2(f)$ is statistically significant at 5% confidence.

Table A.1: Summary Statistics

The table holds summary statistics for the country and region data. ”Co.” is the total number of companies in the sample. ”Av. Co.”, ”Min. Co.”, and ”Max. Co.” are the average, minimum, and maximum number of companies per month respectively. ”Ret.”, ”SD”, ”Skew.”, and ”Kurt.” are average cross-sectional mean, standard deviation, skewness, and excess kurtosis of stock excess returns respectively. ”M.Cap” is the time-series average of the cross-sectional mean market capitalization of the firms in millions of USD. ”First Obs.” is the first date for which the complete Fama and French (2018) six-factor model is available.

	Co.	Av. Co.	Min. Co.	Max. Co.	Ret.	SD	Skew	Kurt	M.Cap	First Obs.
World	78,410	26,862	12,169	38,271	5.4%	61.4%	1.09	5.60	1,535	Jan 1992
Developing	8,060	2,695	250	5,527	21.9%	67.8%	1.19	4.50	621	Mai 1994
ARG	166	62	33	83	12.4%	60.4%	0.95	5.30	704	Apr 1997
BRA	397	114	10	262	16.1%	62.5%	0.70	3.40	1,658	Jan 2000
CHL	287	121	30	180	22.8%	41.1%	1.41	9.20	1,066	Mai 1994
GRC	439	183	24	331	7.2%	63.0%	1.19	5.70	366	Mai 1998
IND	5,799	1,840	73	4,266	24.3%	70.3%	1.23	4.10	471	Apr 2000
MEX	290	91	32	126	9.3%	45.7%	0.70	5.50	2,357	Mai 1993
TUR	682	286	43	528	21.7%	74.7%	1.30	5.60	443	Mai 1998
Emerged	26,705	11,447	2,769	19,716	5.4%	55.0%	1.27	6.70	950	Jan 1992
CHN	5,309	1,740	1	4,999	6.6%	50.1%	1.03	4.60	1,032	Sep 1998
HKG	3,020	1,207	153	2,404	4.6%	64.0%	1.54	8.10	1,222	Jan 1992
IDN	1,035	355	93	832	7.2%	71.1%	1.55	8.60	529	Mai 1993
JPN	6,060	3,442	1,705	3,992	3.6%	47.0%	1.24	7.10	1,238	Jan 1992
KOR	3,788	1,395	125	2,561	1.8%	66.6%	1.20	5.60	496	Feb 1995
MYS	1,495	775	158	999	10.0%	57.5%	1.65	9.80	361	Jan 1992
NZL	311	109	38	136	5.7%	49.7%	0.81	4.40	418	Aug 1993
PER	189	57	5	112	22.4%	47.8%	1.67	12.20	619	Jan 2007
PHL	348	181	18	261	7.8%	61.5%	1.95	11.20	568	Feb 1997
SGP	1,166	498	93	720	11.0%	58.1%	1.41	7.70	673	Jan 1992
THA	1,201	507	210	864	5.7%	53.0%	1.39	8.10	407	Mai 1993
TWN	2,783	1,182	155	2,126	5.7%	49.2%	1.12	5.30	610	Mai 1997
Established	43,645	12,719	9,130	14,613	2.3%	64.2%	0.96	5.20	2,424	Jan 1992
AUS	3,796	1,230	277	1,819	11.1%	74.4%	1.21	5.10	639	Jan 1992
AUT	209	74	50	93	-0.2%	50.6%	0.98	6.30	1,180	Jan 1992
BEL	337	131	59	161	6.1%	46.4%	0.91	7.10	1,895	Jan 1992
CAN	3,159	1,076	484	1,630	19.6%	69.3%	1.11	5.50	1,255	Jan 1992
CHE	512	216	62	268	5.2%	41.2%	0.62	4.90	3,882	Jan 1992
DEU	1,932	708	214	1,018	-1.9%	66.0%	1.14	6.10	1,971	Jan 1992
DNK	418	151	33	216	-0.7%	50.1%	0.93	6.20	1,363	Jan 1992
ESP	505	154	94	245	1.4%	39.0%	1.01	6.90	3,412	Jan 1992
FIN	315	119	24	181	4.9%	43.8%	0.85	4.50	1,361	Jan 1992
FRA	2,034	663	277	848	6.2%	56.0%	1.07	5.60	2,442	Jan 1992
GBR	5,879	1,668	925	2,243	-5.4%	57.7%	0.90	5.50	1,519	Jan 1992
ITA	892	272	181	417	-1.9%	42.8%	0.94	5.00	2,013	Jan 1992
NLD	392	146	101	220	0.9%	42.4%	0.71	5.50	4,067	Jan 1992

Table A.1: *(continued)*

	Co.	Av. Co.	Min. Co.	Max. Co.	Ret.	SD	Skew	Kurt	M.Cap	First Obs.
NOR	740	187	53	335	4.3%	59.6%	0.82	3.90	891	Jan 1992
SWE	1,639	400	70	998	-5.9%	66.5%	1.00	4.00	1,039	Jan 1992
USA	20,886	5,525	4,184	7,999	0.9%	67.6%	0.90	5.00	3,911	Jan 1992

Table A.2: Number of observations per factor

The table holds the number of complete time series observations for monthly international returns in USD of the Fama and French (2018) six-factor model between 1992 and 2023 in 35 countries. An observation is deemed to be complete if no country factor return is missing.

Factor	Obs.
Market	391
Size	184
Value	200
Investment	203
Profitability	179
Momentum	206

Table A.3: Asset pricing tests local vs. regional vs. rest-of-world factor model with size-value sorted portfolios

The table holds the results of asset pricing regressions for local, regional, and rest-of-world Fama and French (2018) six-factor model returns in USD between 1992 and 2023. The regression setup is $r_{c,i} = \alpha_{c,i} + \sum \beta_{c,i} f_c^{L/R/W} + \varepsilon_{c,i}$, for each country, c. $r_{c,i}$ are returns of portfolio i 3×3 double sorted portfolios based on market capitalization and book-to-market ratios. $f_c^{L/R/W}$ contains the returns of a local, regional or rest-of-world factor model. The local model consists of sorted long-short stock portfolios in the respective country. The regional model is a value-weighted average of local factor models in the respective region, excluding the respective country. The rest-of-world model is a value-weighted average of local factor models of all countries outside the respective region. The regression equation is fitted with GMM. Confidence intervals are calculated using Newey and West (1987) adjusted standard errors using 6 lags. The number of observations for the regressions is fixed for the different factor models in each country. I require at least 120 observations for a test portfolio to be included. The "IDs" column holds the number of portfolios for which the asset pricing test is conducted. "R^2" hold the average goodness-of-fit measure of the regressions. "$\bar{\alpha}_c$ holds the average $\alpha_{c,i}$. "F_{GRS}" holds the Gibbons et al. (1989) test statistic for the regression. "Local", "Regional", and "Rest of World" hold the results for the respective factor models. Statistical significance at confidence levels of 10%, 5%, and 1% are indicated with *, **, and *** respectively.

			Local			*Regional*			*Rest of world*		
Country	Region	IDs	R^2	$\bar{\alpha}_c$	F_{GRS}	R^2	$\bar{\alpha}_c$	F_{GRS}	R^2	$\bar{\alpha}_c$	F_{GRS}
AUS	Established	9	93%	1.38%*	1.94**	62%	3.54%**	2.00**	52%	1.46%**	1.93**
AUT		4	80%	2.34%	5.41***	49%	7.52%***	16.33***	37%	5.56%*	7.72***
BEL		9	77%	1.93%	4.35***	56%	3.21%*	4.16***	35%	3.19%	2.46***
CAN		9	89%	1.34%**	2.17**	67%	2.02%***	2.55***	45%	2.89%	1.45
CHE		9	82%	1.45%	0.98	60%	1.65%	1.57	41%	5.48%**	1.74*
DEU		9	88%	2.03%	2.15**	66%	3.32%**	1.56	40%	2.26%	1.85*
DNK		6	78%	1.78%	1.27	53%	2.89%	1.99*	34%	4.66%	3.80***
ESP		9	84%	1.48%*	1.87*	51%	3.01%*	1.87*	32%	2.49%*	2.26**
FIN		8	75%	1.71%	4.11***	56%	3.11%**	5.49***	36%	5.76%	3.74***
FRA		9	90%	1.46%***	2.73***	68%	1.92%***	2.71***	41%	2.86%***	3.49***
GBR		9	92%	1.52%***	2.65***	68%	2.57%	1.02	44%	1.47%	1.13
ITA		9	89%	1.41%	1.16	50%	3.78%	0.91	31%	2.28%	1.66*
NLD		9	81%	2.32%***	2.71***	66%	2.66%***	2.66***	40%	5.26%***	6.13***
NOR		9	84%	1.92%	2.92***	61%	4.63%	2.39**	41%	3.97%	1.95**
SWE		9	88%	1.88%	1.02	64%	1.94%	1.41	40%	7.58%	1.68*
USA		9	95%	0.82%***	3.09***	68%	6.70%***	4.75***	43%	7.21%***	3.63***
Avg.		8	85%	1.67%	2.53	60%	3.40%	3.33	40%	4.02%	2.91

Table A.3: *(continued)*

Country	Region	IDs	*Local*			*Regional*			*Rest of world*		
			R^2	$\bar{\alpha}_c$	F_{GRS}	R^2	$\bar{\alpha}_c$	F_{GRS}	R^2	$\bar{\alpha}_c$	F_{GRS}
CHN	Emerged	9	95%	2.24%***	3.55***	20%	4.79%***	2.63***	11%	5.16%***	3.85***
HKG		9	90%	1.16%***	2.78***	38%	3.26%**	1.89*	34%	3.02%***	3.22***
IDN		9	86%	4.16%**	6.11***	29%	4.21%	2.23**	31%	7.17%	2.24**
JPN		9	95%	1.00%***	2.17**	15%	2.66%***	3.18***	22%	3.16%***	3.51***
KOR		9	89%	2.81%***	2.50***	45%	5.39%**	2.80***	42%	4.95%**	1.70*
MYS		9	92%	1.82%**	2.08**	18%	2.86%*	1.75*	19%	3.56%**	1.90*
NZL		3	87%	3.00%***	7.00***	40%	2.57%	0.36	51%	2.63%	0.41
PER		2	86%	1.76%	1.87	33%	5.21%	2.22	27%	9.05%	2.09
PHL		8	83%	2.79%**	4.97***	33%	4.20%**	4.02***	30%	3.98%	2.42**
SGP		9	90%	1.92%*	2.35**	42%	2.37%	1.67*	42%	2.67%**	2.20**
THA		9	89%	2.58%**	3.66***	31%	2.71%	1.45	33%	4.37%	2.29**
TWN		9	90%	1.01%	1.18	32%	2.91%	2.55***	34%	1.88%	2.27**
Avg.		8	89%	2.19%	3.35	31%	3.59%	2.23	31%	4.30%	2.34
ARG	Developing	3	88%	3.69%	1.63	18%	5.60%	1.76	17%	5.07%	0.97
BRA		9	92%	4.14%**	2.91***	45%	5.84%	2.03**	48%	2.51%	1.06
CHL		9	88%	1.91%***	8.91***	40%	2.73%*	2.65***	37%	3.12%**	3.29***
GRC		3	89%	3.05%	1.56	29%	5.59%	0.58	31%	3.04%	1.95
IND		9	94%	1.52%	2.39**	42%	7.38%	1.68*	37%	8.48%	1.54
MEX		8	85%	2.17%**	2.67***	44%	3.08%	1.23	49%	3.04%	2.25**
TUR		9	92%	4.04%***	7.38***	31%	9.55%***	3.27***	28%	8.52%**	2.94***
Avg.		7	90%	2.93%	3.92	35%	5.68%	1.89	35%	4.82%	2
Avg.		**8**	**88%**	**2.10%**	**3.09**	**45%**	**3.93%**	**2.67**	**36%**	**4.28%**	**2.53**

Table A.4: Asset pricing tests local vs. regional vs. rest-of-world factor model with equally weighted regional and rest-of-world factors

The table holds the results of asset pricing regressions for local, regional, and rest-of-world Fama and French (2018) six-factor model returns between 1992 and 2023. The regression setup is $r_{c,i} = \alpha_{c,i} + \sum \beta_{c,i} f_c^{L/R/W} + \varepsilon_{c,i}$, for each country, c. $r_{c,i}$ are return of industry portfolio i for GICS 10 portfolios. $f_c^{L/R/W}$ contains the returns of a local, regional or rest-of-world factor model. The local model consists of sorted long-short stock portfolios in the respective country. The regional model is an equally weighted average of local factor models in the respective region, excluding the respective country. The rest-of-world model is an equally weighted average of local factor models of all countries outside the respective region. The regression equation is fitted with GMM. Confidence intervals are calculated using Newey and West (1987) adjusted standard errors using 6 lags. The number of observations for the regressions is fixed for the different factor models in each country. I require at least 120 observations for a test portfolio to be included. The "IDs" column holds the number of portfolios for which the asset pricing test is conducted. "R^2" hold the average goodness-of-fit measure of the regressions. "$\bar{\alpha}_c$ holds the average $\alpha_{c,i}$. "F_{GRS}" holds the Gibbons et al. (1989) test statistic for the regression. "Local", "Regional", and "Rest of World" hold the results for the respective factor models. Statistical significance at confidence levels of 10%, 5%, and 1% are indicated with *, **, and *** respectively.

			Local			*Regional*			*Rest of world*		
Country	Region	IDs	R^2	$\bar{\alpha}_c$	F_{GRS}	R^2	$\bar{\alpha}_c$	F_{GRS}	R^2	$\bar{\alpha}_c$	F_{GRS}
AUS	Established	10	78%	1.47%	0.58	60%	3.05%	0.83	59%	4.31%	1.63*
AUT		3	77%	1.15%	1.15	64%	6.93%*	3.75**	48%	5.66%	1.03
BEL		7	70%	1.39%	0.59	63%	3.03%	1.03	49%	3.95%	0.79
CAN		10	65%	2.35%	1.43	54%	4.26%***	3.39***	46%	4.42%	2.02**
CHE		7	76%	2.08%	1.29	68%	2.81%	1.20	51%	3.17%	1.21
DEU		9	73%	4.18%***	3.45***	69%	5.66%***	4.23***	51%	3.10%	0.42
DNK		5	69%	2.48%	0.53	59%	4.20%*	2.93**	43%	4.94%**	2.53**
ESP		6	70%	2.58%	1.43	62%	3.23%	2.18**	45%	5.21%	1.22
FIN		5	68%	4.27%***	4.41***	64%	2.46%	1.18	48%	4.30%	0.72
FRA		10	76%	2.47%	1.27	75%	5.06%***	4.32***	53%	3.58%*	2.56***
GBR		10	74%	2.64%***	2.86***	66%	4.61%***	3.21***	50%	4.30%	1.45
ITA		8	74%	2.11%	0.90	66%	3.65%**	2.88***	47%	4.67%**	3.08***
NLD		5	76%	3.23%*	2.05*	68%	5.27%***	4.48***	48%	4.27%	1.54
NOR		8	72%	4.52%***	3.57***	62%	4.89%	2.36**	52%	5.96%	1.47
SWE		8	72%	1.80%	1.16	66%	4.43%	2.04**	49%	2.92%	3.13***
USA		10	76%	2.06%*	1.71*	60%	5.41%***	4.97***	46%	4.38%***	2.21**
Avg.		8	73%	2.55%	1.77	64%	4.31%	2.81	49%	4.32%	1.69

Table A.4: *(continued)*

Country	Region	IDs	Local R^2	Local $\bar{\alpha}_c$	Local F_{GRS}	Regional R^2	Regional $\bar{\alpha}_c$	Regional F_{GRS}	Rest of world R^2	Rest of world $\bar{\alpha}_c$	Rest of world F_{GRS}
CHN	Emerged	9	86%	3.12%***	3.47***	19%	3.63%*	2.30**	13%	5.76%	1.86*
HKG	Emerged	10	73%	3.81%***	1.90**	51%	3.62%	1.76*	43%	3.29%*	1.57
IDN	Emerged	8	73%	5.32%**	2.31**	38%	5.56%*	1.67	34%	6.47%	0.84
JPN	Emerged	10	68%	3.15%***	2.74***	23%	3.05%**	2.11**	28%	4.17%***	1.84*
KOR	Emerged	10	73%	3.17%**	2.82***	48%	5.43%	1.84*	45%	5.94%**	2.58***
MYS	Emerged	10	79%	1.95%	1.18	45%	2.58%*	1.51	43%	4.13%***	2.38**
NZL	Emerged	6	76%	3.36%**	3.88***	44%	4.41%***	4.54***	48%	6.22%***	5.89***
PER	Emerged	3	82%	3.61%*	2.35*	42%	2.88%	0.43	31%	6.05%	0.99
PHL	Emerged	8	66%	1.74%	0.40	37%	3.46%	0.85	29%	4.93%	0.57
SGP	Emerged	8	79%	3.51%***	4.30***	64%	6.96%***	4.82***	58%	5.28%***	3.56***
THA	Emerged	8	77%	4.99%***	5.11***	47%	5.90%***	3.96***	40%	7.32%***	4.72***
TWN	Emerged	7	80%	1.25%	0.54	39%	2.03%	0.65	36%	2.42%	0.72
Avg.		8	76%	3.25%	2.58	41%	4.13%	2.2	37%	5.16%	2.29
ARG	Developing	3	81%	7.18%	2.24*	16%	2.85%	0.07	14%	3.94%	0.14
BRA	Developing	6	84%	2.67%	1.12	52%	3.41%	1.23	51%	5.49%	0.80
CHL	Developing	6	84%	3.16%***	4.59***	43%	3.86%**	3.09***	44%	5.46%*	3.54***
GRC	Developing	6	82%	2.97%	1.12	30%	8.42%*	2.37**	41%	7.67%***	2.75**
IND	Developing	10	81%	3.19%*	1.66*	34%	7.90%**	1.54	40%	5.45%	1.13
MEX	Developing	5	82%	2.69%	2.32**	47%	2.33%	0.70	56%	2.66%	0.56
TUR	Developing	6	83%	3.58%**	2.88***	33%	2.76%	0.70	30%	9.16%	0.89
Avg.		6	83%	3.63%	2.28	36%	4.50%	1.39	39%	5.69%	1.4
Avg.		**7**	**76%**	**3.01%**	**2.15**	**51%**	**4.29%**	**2.32**	**43%**	**4.88%**	**1.84**

Table A.5: Asset pricing tests local vs. regional vs. rest-of-world factor model with traditional Fama and French (2017) factor formation methodology

The table holds the results of asset pricing regressions for local, regional, and rest-of-world Fama and French (2018) six-factor model returns between 1992 and 2023. The regression setup is $r_{c,i} = \alpha_{c,i} + \sum \beta_{c,i} f_c^{L/R/W} + \varepsilon_{c,i}$, for each country, c. $r_{c,i}$ are return of industry portfolio i for GICS 10 portfolios. $f_c^{L/R/W}$ contains the returns of a local, regional or rest-of-world factor model. The local model is built following the methodology by Fama and French (2017) that creates 2×3 sorted long-short stock portfolios in the respective country. The regional model is built from sorted long-short stock portfolios in the respective region, excluding stocks from the respective country. The rest-of-world model consists of sorted long-short stock portfolios with stocks from all countries outside the respective region. The regression equation is fitted with GMM. Confidence intervals are calculated using Newey and West (1987) adjusted standard errors using 6 lags. The number of observations for the regressions is fixed for the different factor models in each country. I require at least 120 observations for a test portfolio to be included. The "IDs" column holds the number of portfolios for which the asset pricing test is conducted. "R^2" hold the average goodness-of-fit measure of the regressions. "$\bar{\alpha}_c$ holds the average $\alpha_{c,i}$. "F_{GRS}" holds the Gibbons et al. (1989) test statistic for the regression. "Local", "Regional", and "Rest of World" hold the results for the respective factor models. Statistical significance at confidence levels of 10%, 5%, and 1% are indicated with *, **, and *** respectively.

			Local			*Regional*			*Rest of world*		
Country	Region	IDs	R^2	$\bar{\alpha}_c$	F_{GRS}	R^2	$\bar{\alpha}_c$	F_{GRS}	R^2	$\bar{\alpha}_c$	F_{GRS}
AUS	Established	10	76%	4.86%***	4.60***	57%	2.78%*	1.81*	50%	5.66%***	2.39**
BEL	Established	7	64%	2.82%	0.79	57%	4.76%	1.75	43%	6.26%	1.21
CAN	Established	10	62%	2.73%	1.32	59%	2.62%	1.53	40%	4.36%***	1.61
CHE	Established	7	75%	1.87%	1.20	64%	3.13%	1.06	44%	5.66%	0.77
DEU	Established	9	66%	3.13%	1.16	61%	4.17%	1.04	39%	3.18%	1.17
FRA	Established	10	75%	2.38%	2.56***	67%	3.97%	3.22***	44%	6.91%	2.35**
GBR	Established	10	67%	2.00%	0.71	59%	3.79%	1.27	40%	2.90%	1.19
ITA	Established	8	70%	3.41%	1.35	56%	5.09%*	1.58	38%	3.04%	1.05
NLD	Established	5	78%	2.97%	0.88	71%	7.12%**	2.47**	54%	4.64%**	2.66**
NOR	Established	8	64%	4.93%*	2.11**	53%	12.68%***	4.90***	32%	9.71%***	4.84***
SWE	Established	8	71%	3.94%	1.62	65%	5.62%*	2.35**	47%	13.70%**	2.97***
USA	Established	10	69%	1.65%	1.28	56%	4.47%***	2.20**	38%	7.69%***	2.23**
Avg.		8	70%	3.06%	1.63	60%	5.02%	2.1	42%	6.14%	2.04
CHN	Emerged	9	86%	2.41%*	1.81*	20%	3.24%	1.26	13%	7.79%	0.90
HKG	Emerged	10	70%	3.08%***	2.54***	51%	7.52%***	3.22***	39%	2.80%	1.42
IDN	Emerged	8	65%	8.03%***	3.06***	21%	5.35%	1.38	25%	9.69%	1.35
JPN	Emerged	10	64%	1.26%	0.61	17%	2.45%	0.95	27%	1.79%	0.75
KOR	Emerged	10	72%	2.77%	1.18	42%	8.17%	1.21	47%	3.98%	1.72*
MYS	Emerged	10	75%	3.87%***	3.23***	34%	2.79%**	1.55	37%	3.10%	1.52
SGP	Emerged	8	76%	6.16%***	12.15***	48%	4.57%***	2.93***	53%	5.18%***	3.59***
THA	Emerged	8	75%	5.34%***	4.22***	32%	8.40%***	3.33***	36%	5.68%***	4.01***
TWN	Emerged	7	77%	2.60%***	2.20**	43%	7.38%***	2.48**	42%	5.13%*	1.39
Avg.		9	73%	3.95%	3.44	34%	5.54%	2.04	35%	5.02%	1.85
Avg.		**9**	**71%**	**3.44%**	**2.41**	**49%**	**5.24%**	**2.07**	**39%**	**5.66%**	**1.96**

Table A.6: Asset pricing tests local vs. regional vs. rest-of-world factor model with sorting aggregation procedure

The table holds the results of asset pricing regressions for local, regional, and rest-of-world Fama and French (2018) six-factor model returns between 1992 and 2023. The regression setup is $r_{c,i} = \alpha_{c,i} + \sum \beta_{c,i} f_c^{L/R/W} + \varepsilon_{c,i}$, for each country, c. $r_{c,i}$ are return of industry portfolio i for GICS 10 portfolios. $f_c^{L/R/W}$ contains the returns of a local, regional or rest-of-world factor model. The local model consists of sorted long-short stock portfolios in the respective country. The regional model is built from sorted long-short stock portfolios in the respective region, excluding stocks from the respective country. The rest-of-world model consists of sorted long-short stock portfolios with stocks from all countries outside the respective region. The regression equation is fitted with GMM. Confidence intervals are calculated using Newey and West (1987) adjusted standard errors using 6 lags. The number of observations for the regressions is fixed for the different factor models in each country. I require at least 120 observations for a test portfolio to be included. The "IDs" column holds the number of portfolios for which the asset pricing test is conducted. "R^2" hold the average goodness-of-fit measure of the regressions. "$\bar{\alpha}_c$ holds the average $\alpha_{c,i}$. "F_{GRS}" holds the Gibbons et al. (1989) test statistic for the regression. "Local", "Regional", and "Rest of World" hold the results for the respective factor models. Statistical significance at confidence levels of 10%, 5%, and 1% are indicated with *, **, and *** respectively.

			Local			*Regional*			*Rest of world*		
Country	Region	IDs	R^2	$\bar{\alpha}_c$	F_{GRS}	R^2	$\bar{\alpha}_c$	F_{GRS}	R^2	$\bar{\alpha}_c$	F_{GRS}
AUS	Established	10	78%	1.47%	0.58	59%	1.40%	0.55	53%	1.57%	0.53
AUT		3	77%	1.15%	1.15	54%	4.93%	1.83	40%	2.97%	0.96
BEL		7	70%	1.39%	0.59	57%	1.77%	0.31	41%	2.82%	0.66
CAN		10	65%	2.35%	1.43	58%	2.42%	1.41	39%	4.82%	1.05
CHE		7	76%	2.08%	1.29	64%	2.44%	0.56	48%	6.08%	1.92*
DEU		9	73%	4.18%***	3.45***	64%	3.94%	2.67***	45%	2.93%	0.88
DNK		5	69%	2.48%	0.53	54%	3.66%	1.31	39%	8.08%**	2.93**
ESP		6	70%	2.58%	1.43	54%	1.26%	0.48	40%	2.62%	1.42
FIN		5	68%	4.27%***	4.41***	58%	3.05%*	2.41**	42%	4.19%	1.17
FRA		10	76%	2.47%	1.27	70%	3.24%	3.30***	46%	3.64%	0.78
GBR		10	74%	2.64%***	2.86***	64%	2.45%	1.25	45%	1.86%	0.95
ITA		8	74%	2.11%	0.90	59%	2.42%	1.55	41%	3.92%	2.08**
NLD		5	76%	3.23%*	2.05*	64%	2.61%*	2.47**	43%	4.23%	1.45
NOR		8	72%	4.52%***	3.57***	59%	4.35%	2.23**	45%	2.61%	0.62
SWE		8	72%	1.80%	1.16	63%	2.50%	0.71	43%	5.75%	2.02**
USA		10	76%	2.06%*	1.71*	62%	4.56%***	5.03***	41%	5.64%***	2.37**
Avg.		8	73%	2.55%	1.77	60%	2.94%	1.76	43%	3.98%	1.36

Table A.6: *(continued)*

Country	Region	IDs	Local R^2	Local $\bar{\alpha}_c$	Local F_{GRS}	Regional R^2	Regional $\bar{\alpha}_c$	Regional F_{GRS}	Rest of world R^2	Rest of world $\bar{\alpha}_c$	Rest of world F_{GRS}
CHN	Emerged	9	86%	3.12%***	3.47***	24%	3.42%**	2.27**	12%	5.77%	1.81*
HKG		10	73%	3.81%***	1.90**	51%	2.68%	1.64*	40%	3.00%*	1.70*
IDN		8	73%	5.32%**	2.31**	27%	6.92%	1.85*	30%	5.78%	1.97*
JPN		10	68%	3.15%***	2.74***	21%	2.74%	1.24	28%	1.48%	0.69
KOR		10	73%	3.17%**	2.82***	42%	3.58%	0.96	44%	3.52%	1.43
MYS		10	79%	1.95%	1.18	34%	2.51%	1.16	37%	2.37%	1.32
NZL		6	76%	3.36%**	3.88***	39%	5.56%**	4.01***	49%	2.69%	2.02*
PER		3	82%	3.61%*	2.35*	30%	7.61%	1.49	31%	10.50%	1.44
PHL		8	66%	1.74%	0.40	26%	4.41%	0.97	27%	2.42%	0.57
SGP		8	79%	3.51%***	4.30***	49%	3.33%***	3.29***	53%	3.36%***	4.10***
THA		8	77%	4.99%***	5.11***	36%	7.81%***	3.75***	37%	4.91%***	3.91***
TWN		7	80%	1.25%	0.54	29%	3.45%	0.96	32%	2.93%	0.96
Avg.		8	76%	3.25%	2.58	34%	4.50%	1.97	35%	4.06%	1.83
ARG	Developing	3	81%	7.18%	2.24*	17%	7.26%	0.73	12%	3.98%	0.20
BRA		6	84%	2.67%	1.12	48%	7.09%	1.45	46%	3.49%	0.70
CHL		6	84%	3.16%***	4.59***	42%	3.49%**	2.85**	37%	2.40%	2.45**
GRC		6	82%	2.97%	1.12	30%	7.46%**	2.58**	38%	6.06%*	1.81*
IND		10	81%	3.19%*	1.66*	35%	8.75%**	1.87**	34%	8.45%	1.35
MEX		5	82%	2.69%	2.32**	45%	2.03%	0.66	52%	1.78%	0.20
TUR		6	83%	3.58%**	2.88***	30%	3.09%	0.69	30%	11.11%	0.90
Avg.		6	83%	3.63%	2.28	35%	5.60%	1.55	36%	5.33%	1.09
Avg.		**7**	**76%**	**3.01%**	**2.15**	**46%**	**4.00%**	**1.79**	**39%**	**4.28%**	**1.47**

Table A.7: Asset pricing tests model comparison for local vs. regional vs. rest-of-world factor model

The table holds the regionally aggregated results of asset pricing regressions for local, regional, and rest-of-world factor model returns in USD between 1992 and 2023. The regression setup is $r_{c,i} = \alpha_{c,i} + \sum \beta_{c,i} f_c^{L/R/W} + \varepsilon_{c,i}$, for each country, c. $r_{c,i}$ are return of industry portfolio i for GICS 10 portfolios. $f_c^{L/R/W}$ contains the returns of a local, regional or rest-of-world factor model. The local model consists of sorted long-short stock portfolios in the respective country. The regional model is a value-weighted average of local factor models in the respective region, excluding the respective country. The rest-of-world model is a value-weighted average of local factor models of all countries outside the respective region. The models are evaluated once using their own market factor (columns 3-6) and once where the market factor is fixed to be the market factor of the local model for the regional and rest-of-world models (columns 7-10). The regression equation is fitted using GMM. Confidence intervals are calculated using Newey and West (1987) adjusted standard errors with 6 lags. The number of observations for the regressions is fixed for the different factor models in each country. I require at least 120 observations for a test portfolio to be included. The "Region" column shows over what region the results are aggregated. The "Models" column shows whether the aggregated results are for a local, regional, or rest-of-world model. "R^2" hold the average goodness-of-fit measure of the regressions. "$\bar{\alpha}_c$ holds the average $\alpha_{c,i}$. "F_{GRS}" holds the Gibbons et al. (1989) test statistic for the regression. "ΔSh^2" holds the results of the Barillas and Shanken (2017) squared Sharpe ratio test. Results in this column are always evaluated against the local model. E.g., for the regional model, the column holds the result from $Sh^2(f^R) - Sh^2(f^L)$. "Local", "Regional", and "Rest of World" hold the results for the respective factor models. The factor models are "FF6$_c$", the Fama and French (2018) six-factor model with a cash profitability factor, "FF5", the Fama and French (2015) five-factor model, "HXZ4", the Hou et al. (2015) q-factor model, and "FF3", the Fama and French (1993) three-factor model.

		Model Market Factor				*Local Market Factor*			
Region	Model	R^2	$\bar{\alpha}_c$	F_{GRS}	ΔSh^2	R^2	$\bar{\alpha}_c$	F_{GRS}	ΔSh^2
					FF6$_c$				
Developing	Local	90%	2.93%	3.92		83%	3.66%	2.37	
	Regional	35%	5.68%	1.89	0.19	78%	4.05%	1.87	0.07
	Rest of World	35%	4.82%	2.00	0.24	78%	4.34%	1.91	0.27
Established	Local	85%	1.67%	2.53		73%	2.54%	1.75	
	Regional	60%	3.40%	3.33	0.22	70%	2.95%	1.75	0.24
	Rest of World	40%	4.02%	2.91	0.27	68%	3.14%	1.63	0.16
Emerged	Local	89%	2.19%	3.35		76%	3.06%	2.46	
	Regional	31%	3.59%	2.23	0.13	71%	3.53%	1.97	0.16
	Rest of World	31%	4.30%	2.34	0.38	72%	3.04%	1.95	0.31
Avg.	**Local**	**88%**	**2.10%**	**3.09**		**76%**	**2.95%**	**2.12**	
	Regional	**45%**	**3.93%**	**2.67**	**0.17**	**72%**	**3.37%**	**1.85**	**0.17**
	Rest of World	**36%**	**4.28%**	**2.53**	**0.32**	**71%**	**3.35%**	**1.80**	**0.25**
					FF5				
Developing	Local	82%	3.77%	2.34		82%	3.77%	2.34	
	Regional	34%	5.87%	1.77	0.08	78%	4.75%	2.33	0.06
	Rest of World	35%	4.85%	0.98	0.07	78%	4.22%	1.84	0.05

Table A.7: *(continued)*

Region	Model	Model Market Factor R^2	$\bar{\alpha}_c$	F_{GRS}	ΔSh^2	Local Market Factor R^2	$\bar{\alpha}_c$	F_{GRS}	ΔSh^2
Established	Local	72%	2.37%	1.48		72%	2.37%	1.48	
	Regional	59%	2.80%	1.38	0.05	70%	2.72%	1.50	0.05
	Rest of World	42%	3.04%	1.50	0.09	67%	2.60%	1.63	0.04
Emerged	Local	76%	3.22%	2.86		76%	3.22%	2.86	
	Regional	32%	3.89%	2.20	0.07	71%	3.50%	2.52	0.05
	Rest of World	35%	3.75%	1.75	0.09	71%	2.88%	2.16	0.08
Avg.	**Local**	**75%**	**2.94%**	**2.13**		**75%**	**2.94%**	**2.13**	
	Regional	**45%**	**3.79%**	**1.74**	**0.07**	**72%**	**3.39%**	**2.02**	**0.05**
	Rest of World	**38%**	**3.65%**	**1.48**	**0.09**	**71%**	**3.02%**	**1.85**	**0.06**
					HXZ4				
Developing	Local	81%	3.41%	1.58		81%	3.41%	1.58	
	Regional	37%	5.95%	1.54	0.08	78%	4.86%	2.38	0.03
	Rest of World	36%	4.63%	1.09	0.11	78%	4.69%	1.94	0.06
Established	Local	73%	2.35%	1.52		73%	2.35%	1.52	
	Regional	61%	3.84%	1.88	0.09	71%	3.13%	1.81	0.08
	Rest of World	44%	4.22%	1.51	0.07	69%	2.76%	1.39	0.04
Emerged	Local	76%	2.85%	2.23		76%	2.85%	2.23	
	Regional	36%	5.18%	1.65	0.05	72%	3.02%	1.78	0.03
	Rest of World	38%	4.94%	1.92	0.10	72%	3.02%	1.97	0.09
Avg.	**Local**	**75%**	**2.73%**	**1.77**		**75%**	**2.73%**	**1.77**	
	Regional	**48%**	**4.72%**	**1.74**	**0.07**	**73%**	**3.44%**	**1.92**	**0.05**
	Rest of World	**41%**	**4.55%**	**1.57**	**0.09**	**72%**	**3.24%**	**1.70**	**0.07**
					FF3				
Developing	Local	82%	3.91%	2.50		82%	3.91%	2.50	
	Regional	33%	4.75%	1.42	0.00	77%	3.98%	2.24	0.00
	Rest of World	35%	4.11%	1.33	-0.01	77%	3.80%	1.86	-0.01
Established	Local	70%	2.27%	1.47		70%	2.27%	1.47	
	Regional	57%	2.93%	1.49	0.03	68%	2.56%	1.40	0.03
	Rest of World	42%	3.49%	1.76	-0.01	67%	2.75%	1.84	-0.01
Emerged	Local	74%	3.13%	2.84		74%	3.13%	2.84	
	Regional	31%	4.07%	2.25	0.00	71%	3.29%	2.50	-0.01
	Rest of World	34%	3.69%	1.92	0.03	71%	2.93%	2.26	0.04
Avg.	**Local**	**74%**	**2.89%**	**2.14**		**74%**	**2.89%**	**2.14**	
	Regional	**43%**	**3.68%**	**1.74**	**0.01**	**71%**	**3.10%**	**1.94**	**0.01**
	Rest of World	**38%**	**3.69%**	**1.73**	**0.01**	**70%**	**3.02%**	**1.99**	**0.01**

Table A.8: Asset pricing tests local vs. regional vs. rest-of-world factor model with local market factor in every model

The table holds the results of asset pricing regressions for local, regional, and rest-of-world Fama and French (2018) six-factor model returns between 1992 and 2023. The regression setup is $r_{c,i} = \alpha_{c,i} + \sum \beta_{c,i} f_c^{L/R/W} + \varepsilon_{c,i}$, for each country, c. $r_{c,i}$ are return of industry portfolio i for GICS 10 portfolios. $f_c^{L/R/W}$ contains the returns of a local, regional or rest-of-world factor model. The local model consists of sorted long-short stock portfolios in the respective country. The regional model is a value-weighted average of local factor models in the respective region, excluding the respective country. The rest-of-world model is a value-weighted average of local factor models of all countries outside the respective region. The market factor is fixed to be the market factor of the local model for all other models. The regression equation is fitted with GMM. Confidence intervals are calculated using Newey and West (1987) adjusted standard errors using 6 lags. The number of observations for the regressions is fixed for the different factor models in each country. I require at least 120 observations for a test portfolio to be included. The "IDs" column holds the number of portfolios for which the asset pricing test is conducted. "R^2" hold the average goodness-of-fit measure of the regressions. "$\bar{\alpha}_c$ holds the average $\alpha_{c,i}$. "F_{GRS}" holds the Gibbons et al. (1989) test statistic for the regression. "Local", "Regional", and "Rest of World" hold the results for the respective factor models. Statistical significance at confidence levels of 10%, 5%, and 1% are indicated with *, **, and *** respectively.

			Local			*Regional*			*Rest of world*		
Country	Region	IDs	R^2	$\bar{\alpha}_c$	F_{GRS}	R^2	$\bar{\alpha}_c$	F_{GRS}	R^2	$\bar{\alpha}_c$	F_{GRS}
AUS	Established	10	78%	1.47%	0.58	78%	1.66%	0.88	77%	2.22%	1.16
AUT		3	77%	1.15%	1.15	71%	4.16%	1.97	71%	2.14%	0.90
BEL		7	70%	1.39%	0.59	66%	2.11%	0.67	64%	1.70%	0.57
CAN		10	65%	2.35%	1.43	66%	2.63%	1.54	59%	3.68%*	1.57
CHE		7	76%	2.08%	1.29	75%	2.09%	0.97	72%	2.76%	1.22
DEU		9	73%	4.18%***	3.45***	71%	2.82%	1.40	69%	2.95%	1.18
DNK		5	69%	2.48%	0.53	65%	2.62%	0.57	63%	4.52%	1.40
ESP		6	70%	2.58%	1.43	67%	1.64%	1.15	65%	2.71%**	2.71**
FIN		5	68%	4.27%***	4.41***	60%	4.66%**	3.12***	59%	2.53%	0.48
FRA		10	76%	2.47%	1.27	78%	3.08%	2.82***	74%	2.70%	1.58
GBR		10	74%	2.64%***	2.86***	73%	2.04%	1.12	69%	2.94%*	1.74*
ITA		8	74%	2.11%	0.90	72%	3.27%	1.92*	70%	4.13%	2.27**
NLD		5	76%	3.23%*	2.05*	73%	3.35%***	3.29***	70%	3.60%	1.76
NOR		8	72%	4.52%***	3.57***	69%	3.83%**	2.82***	67%	3.70%**	3.15***
SWE		8	72%	1.80%	1.16	72%	2.33%	0.56	69%	2.21%	1.56
USA		10	76%	2.06%*	1.71*	71%	4.11%***	4.04***	66%	2.76%**	2.42***
Avg.		8	73%	2.55%	1.77	70%	2.90%	1.8	68%	2.95%	1.6

Table A.8: *(continued)*

Country	Region	IDs	Local R^2	Local $\bar{\alpha}_c$	Local F_{GRS}	Regional R^2	Regional $\bar{\alpha}_c$	Regional F_{GRS}	Rest of world R^2	Rest of world $\bar{\alpha}_c$	Rest of world F_{GRS}
CHN	Emerged	9	86%	3.12%***	3.47***	82%	3.69%**	2.64***	81%	2.38%**	2.64***
HKG	Emerged	10	73%	3.81%***	1.90**	68%	2.69%	1.61	69%	2.45%	1.46
IDN	Emerged	8	73%	5.32%**	2.31**	69%	4.49%*	1.83*	69%	3.87%	1.56
JPN	Emerged	10	68%	3.15%***	2.74***	62%	2.18%	0.82	65%	1.78%	0.95
KOR	Emerged	10	73%	3.17%**	2.82***	70%	2.73%	1.01	70%	2.35%	1.14
MYS	Emerged	10	79%	1.95%	1.18	74%	3.18%*	1.73*	75%	2.33%	1.44
NZL	Emerged	6	76%	3.36%**	3.88***	71%	4.80%***	5.64***	72%	3.02%*	2.76**
PER	Emerged	3	82%	3.61%*	2.35*	78%	3.23%	1.49	79%	4.46%**	3.10**
PHL	Emerged	8	66%	1.74%	0.40	61%	2.87%	0.57	62%	1.84%	0.56
SGP	Emerged	8	79%	3.51%***	4.30***	76%	4.50%***	3.36***	77%	4.25%***	4.55***
THA	Emerged	8	77%	4.99%***	5.11***	72%	5.49%***	4.25***	73%	5.22%***	5.19***
TWN	Emerged	7	80%	1.25%	0.54	72%	1.84%	0.57	72%	1.32%	0.49
Avg.		8	76%	3.25%	2.58	71%	3.47%	2.13	72%	2.94%	2.15
ARG	Developing	3	81%	7.18%	2.24*	72%	3.86%	0.62	72%	6.28%	2.34*
BRA	Developing	6	84%	2.67%	1.12	83%	3.21%	1.08	83%	2.98%	0.50
CHL	Developing	6	84%	3.16%***	4.59***	83%	3.40%***	4.58***	83%	2.41%**	2.57**
GRC	Developing	6	82%	2.97%	1.12	71%	10.08%*	2.31**	70%	8.05%	1.72
IND	Developing	10	81%	3.19%*	1.66*	77%	3.83%	1.67*	78%	3.40%	1.31
MEX	Developing	5	82%	2.69%	2.32**	81%	3.30%	2.51**	81%	3.67%*	2.45**
TUR	Developing	6	83%	3.58%**	2.88***	80%	4.47%*	2.53**	80%	3.52%	1.70
Avg.		6	83%	3.63%	2.28	78%	4.59%	2.18	78%	4.33%	1.8
Avg.		**7**	**76%**	**3.01%**	**2.15**	**72%**	**3.44%**	**1.99**	**71%**	**3.22%**	**1.83**

B Detailed Variable Description

This section provides a detailed description of the variables used in the factor models for the analysis. In the construction methodology, I follow (Jensen et al., 2023). Capitalized items refer to the respective fields in the COMPUSTAT and CRSP library. It is assumed that accounting variables become available 4 months after the end of the accounting period. The accounting data is sourced from the quarterly and yearly COMPUSTAT files. To make quarterly income and cash flow items comparable to the corresponding annual items, the sum of the last four quarters is taken. Quarterly and annual data is created separately and the most recent characteristic is taken from each data set to create a final joint data set. A suffix of "*" indicates that the alteration for missing values was provided earlier in the text.

The sorting variables for the five long-short portfolios of the Fama and French (2018) six-factor model are defined as follows:

Size is the market equity in million USD of the stock that is computed as the shares outstanding times the price ($SHROUT/1000 * |PRC|$).

Value is defined as the book equity to market equity ratio of the stock. Book equity is defined as the sum of shareholder equity and deferred taxes and investment credit minus preferred stock ($SEQ + TXDITC - PSTK$). If $TXDITC$ or $PSTK$ are missing, they are set to zero.

Investment measures the asset growth of total assets (AT) over the last 12 months as $\frac{AT_t}{AT_{t-12}} - 1$. If it is not available I use the sum of shareholder equity, long-term debt, current liabilities, liabilities, deferred taxes and investment tax credit ($SEQ + DLTT + LCT + LO + TXDITC$). If LCT, LO, or $TXDITC$ are missing, they are set to zero.

Profitability is defined as the operating profitability to book equity* ratio. Operating profitability is calculated as earnings before interest, taxes, depreciation and amortization minus interest rate expenses ($EBITDA - XINT$). If $EBITDA$ is missing, it is replaced by operating income before depreciation ($OIBDP$). If this is unavailable, sales minus operating expenses is used ($SALE - XOPR$)[23]. If this is missing, gross profit minus selling, general and administrative expenses ($GP - XSGA$)[24] is used.

Momentum is defined as the cumulative return of the past 12 months while skipping the most recent month.

Additional sorting variables for the other factor models used as robustness checks are defined as follows:

Cash Profitability is defined as the ratio between cash profitability to 12-month lagged total assets (AT^*_{t-12}). Cash profitability is defined as the sum of $EBITDA^*$ and research and development expenses (XRD) minus operating accruals where XRD is set to zero if it is missing. Operating accruals is defined as income before extraordinary items minus operating activities ($IB - OANCF$)[25]. If that is unavailable, the sum of the yearly change in current

[23] If $SALE$ is unavailable, total revenue ($REVT$) is used. If $XOPR$ is missing, the sum of cost of goods sold and selling, general and administrative expenses ($COGS + XSGA$) is used.

[24] If GP is unavailable, it is defined as sales minus cost of goods sold ($SALE - COGS$). If this is unavailable, it is defined as total revenue minus cost of goods sold ($REVT - COGS$).

[25] If IB is missing, net income minus extraordinary items and discontinued operations is used ($NI - XIDO$). If $XIDO$ is missing, the sum of extraordinary items and discontinued operations ($XI + DO$) is used. If DO is unavailable, it is set to zero. If NI is unavailable, IB is defined as earnings before tax and extraordinary

operating working capital[26] and the the yearly change in net non-current operating assets[27] is used.

Profitability following Hou et al. (2015) is defined as the ratio between quarterly income (IBY) and the book equity* lagged by one quarter.

C Technical Details Joint Analysis

This section contains the technical details for the methods employed in Section 3.2.

Step 1: Joint Decomposition: The iPCA algorithm by Tang and Allen (2021) allows to perform eigendecomposition of multiple data sets simultaneously.

Suppose there are M coupled data matrices $Z_1, \ldots, Z_M$ of dimensions $n \times p_1$, where n is the number of samples and p_m is the number of features in Z_m. Let $p := \sum_{m=1}^{M} p_m$ and $\tilde{Z} := |Z_1, \ldots, Z_M|$. Under the iPCA model, it is assumed that each Z_m arises from a matrix variate normal distribution,

$$Z_m \sim N_{n,p_k}(0, \Sigma \otimes \Delta_m) \qquad (m = 1, \ldots, M) \tag{C.1}$$

where Σ is an $n \times n$ covariance matrix that is jointly shared by all data matrices, and Δ_m is a $p_m \times p_m$ column covariance matrix that is specific to Z_m. The iPCA model aims to maximize

items minus the sum of income taxes and non-controlling interest ($PI - TXT - MII$) where MII is set to zero if missing. If PI is missing, earnings before interest and taxes minus interest and related expense plus the sum of special items and non-operating income ($EBIT - XINT + SPI + NOPI$) is taken where SPI and $NOPI$ are set to zero if missing. If $EBIT$ is missing, operating income after depreciation $OIADP$ is used. If this is unavailable, $EBITDA^*$ minus depreciation and amortization (DP) is used.

[26] This is defined as current operating assets minus current operating liabilities. Current operating assets is defined as current assets minus cash and short-term investmens ($ACT - CHE$). If ACT is missing, the sum of receivables, inventories, cash and short-term inventories and other current assets ($RECT + INVT + CHE + ACO$) is used. Current operating liabilities is defined as current liabilities minus debt in current liabilities ($LCT - DLC$) where DLC is set to zero if missing. If LCT is unavailable, the sum of accounts payable, debt in current liabilities, income taxes payable, and current liabilities ($AP + DLC + TXP + LCO$) is used.

[27] This is defnied as non-current operating assets minus non-current operating liabilities. Non-current operating assets are total assets minus current assets minus investments and advances ($AT^* - ACT - IVAO$). If ACT is missing, the sum of receivables, inventories, cash and short-term inventories and other current assets ($RECT + INVT + CHE + ACO$) is used. Non-current operating liabilities are defined as total liabilities minus current liabilities minus long term debt ($LT - LCT - DLTT$). If LCT is unavailable, the sum of accounts payable, debt in current liabilities, income taxes payable, and current liabilities ($AP + DLC + TXP + LCO$) is used.

both, Σ and $\Delta_1, \ldots, \Delta_M$ simultaneously. This is done in a penalized maximum likelihood framework as

$$\begin{aligned}\hat{\Sigma}^{-1}, \hat{\Delta}_1^{-1}, \ldots, \hat{\Delta}_M^{-1} = \max_{\substack{\Sigma \succ 0 \\ \Delta_1^{-1}, \ldots, \Delta_M \succ 0}} \Bigg\{ & p \log |\Sigma^{-1}| + n \sum_{m=1}^{M} log|\Delta_m^{-1}| \\ & - 1 \sum_{m=1}^{M} \mathrm{tr}(\Sigma^{-1} Z_m \Delta_m^{-1} Z_m^T) \\ & - P(\Sigma^{-1}, \Delta_1^{-1}, \ldots, \Delta_M^{-1}) \Bigg\},\end{aligned} \tag{C.2}$$

where $\hat{\Sigma}^{-1}$ and $\hat{\Delta}_m^{-1}$ are sample estimates of Σ^{-1} and Δ_m^{-1} and $P^*(\Sigma^{-1}, \Delta^{-1}) = \sum_{m=1}^{M} \lambda_m ||\Sigma^{-1} \otimes \Delta_m^{-1}||_F^2||$ is a multiplicative Frobenius penalty for which λ_m is a penalty term. I follow Tang and Allen (2021) and estimate the respective penalty terms in their proposed ”flip-flop” algorithm. Lastly, since I am mostly interested in the joint patterns, the eigenvectors can be retrieved by performing an eigendecomposition on Σ.

Step 2: Dimension Reduction: The parallel analysis method by Horn (1965) applies the Kaiser (1960) criterion for eigenvalue retention and adjusts eigenvalues for random errors using bootstrapping.

Consider a correlation matrix, $\mathbb{C}_r$, for independent random variables. In theory, $\mathbb{C}_r = \mathbb{I}$ where $\mathbb{I}$ is an identity matrix of same dimensions as $\mathbb{C}_r$. Therefore, the PCs should be the same as the original variables, and all eigenvalues, d_i, should be equal to 1. The Kaiser (1960) rule states that because of this only those PCs whose $d_i \geq 1$ should be retained. Horn (1965) proposes a procedure that is based on the Kaiser (1960) rule but corrects for spurious correlation that can be attributed to sampling error. The procedure corrects eigenvalues of the original data, d_i^H by subtracting the mean value of the randomly generated eigenvectors $\bar{d}_i^r$, i.e. $d_i^H = d_i - \bar{d}_i^r$, where d_i^H is the eigenvalue after the Horn (1965) correction.

Some authors suggest taking a higher percentile (e.g., the 95^{th}) of the distribution of the randomly generated eigenvalues (e.g., Glorfeld, 1995) to better correct for spurious correlation. However, others argue that the cutoff of the Kaiser (1960) criterion is too strict, and thus too few PCs are contained (e.g., Jolliffe, 1972, 1973). Because of this divergence in the literature, I stick to Horn (1965) original approach and subtract the mean of the randomly generated eigenvalues.

Step 3: Assess Clustering Tendency: The clustering statistic by Hopkins (1954) measures how much the data is spread out compared to randomly created data in the same span. It can be calculated as follows. Let V be a space composed of e eigenvectors. Let Y be a set of m data points that are placed uniformly randomly in the e-dimensional sample space. Define u_j as the Euclidean distance from y_k to its nearest neighbor in V and w_k as the Euclidean distance from a randomly selected data point in V to its nearest neighbor (m out of the available n points are drawn at random for this purpose). The Hopkins (1954) statistic is defined as

$$H = \frac{\sum_{j=1}^{m} u_j^e}{\sum_{j=1}^{m} u_j^e + \sum_{j=1}^{m} w_j^e}. \tag{C.3}$$

The statistic compares the nearest-neighbor distribution of randomly created data points to that of randomly selected data points. If the data was randomly distributed, H should be about 0.5. If the data is clustered, the distance between randomly created data points and their nearest neighbor should be larger than that of randomly selected data points and their nearest neighbor, so H should be larger than 0.5. A value of $H > 0.7$ is deemed to indicate clustering (Banerjee & Davé, 2004).

It is recommended to set $m \ll n$, for example, $m < 0.1n$. With such a condition, it can be ensured that all $2m$ draws are statistically independent and H follows a beta distribution with shape parameters (m, m). However, it is also suggested to ensure $m > 10$ to avoid small-sample problems Cross and Jain (1982). Because the eigenvectors from my factor have only 35 entries (one for each of the 35 countries in my sample) I either disregard the recommendation to set $m < 0.1n$ or $m > 10$. To fix this issue I run the procedure for $m = 10$ and $m = 3$.

I estimate H using 10.000 bootstrapped draws. To meet both of the above requirements I repeat the simulation two times, for $m = 10$ and $m = 3$. Both setups lead to similar results, indicating that the data is clustered ($\bar{H}_{m=10} = 0.76, \bar{H}_{m=3} = 0.73$). For the $m = 10$ setup, the result is statistically different from 0.5 at 10% confidence.

Step 4: Optimal Number of Clusters: To identify the optimal number of clusters, I use three different metrics, the within sum of squares (WSS), the silhouette measure, and the gap statistic. For all measures, I use the Euclidean distance as the measure of distance between points and define the clusters using the k-means clustering algorithm.

The WSS is the average distance between points within a cluster as:

$$WSS = \sum_{o=1}^{O} \sum_{x \in X_i} D(x, C_i)^2, \tag{C.4}$$

where O is the total number of clusters investigated, $D(x, y)$ is the euclidean distance between x and y, X_i is a cluster of data points x, and $C_i = \frac{1}{|X_i|} \sum_{x \in X_i} x$ is the centroid of cluster X_i. The optimum number of clusters is the point where the WSS cannot be improved strongly for one more cluster (elbow method).

The silhouette method by Rousseeuw (1987) compares the distance of each data point to all other points within its cluster and compares it to the average distance of this point to all other points outside its cluster. The silhouette value of data point x in cluster X_i is defined as

$$s_x = \frac{b_x - a_x}{\max\{a_x, b_x\}}, \tag{C.5}$$

where $a_x = \frac{1}{X_i - 1}{}_i \sum_{x \in X_i, x \neq y} D(x, y)$ is the the mean distance between the data point x and all other points in its cluster and $b_x = \min_{i \neq j} \frac{1}{X_j} \sum_{y \in X_j} D(x, y)$ is the mean distance of x to all other points in any other cluster. The optimal number of clusters maximizes the average silhouette value of all data points.

The gap statistic by Tibshirani et al. (2001) compares the average intra-cluster variation of the actual clusters to the intra-cluster variation of uniformly randomly generated data in the same range of observed values. It is calculated as

$$g_x = \mathbb{E}^*\{\log(a_x)\} - \log(a_x)) \tag{C.6}$$

where $\mathbb{E}^*$ denotes the expectation under the reference distribution. It is defined via bootstrapping and defines a reference dataset under the assumption of randomly distributed data. The reference dataset X_i^* has the same number of observations as X_i which are drawn uniformly in the interval $[\min(x), \max(x)]$. This reference should not be exposed to clustering and thus the optimal number of clusters is the one that maximizes the gap statistic. For the analysis at hand, I calculate the gap statistic using 10,000 bootstrapped samples.

To validate the results of the above analysis for the choice of the clustering algorithm, I calculate the above summary statistics again using the Partitioning Around Medoids (PAM) algorithm (Kaufman & Rousseeuw, 1990) instead of the k-means clustering. The results are illustrated in Figure A.4 and look almost identical to those where the k-means clustering is used.

D Bayesian Model Selection Algorithm Specifications

Recall , that the log marginal likelihood of $\mathbb{M}_j (j \neq J)$ with sample data y can be calculated in closed form as

$$\log \tilde{m}(y|\mathbb{M}_j) = \log \tilde{m}(\tilde{f}|\mathbb{M}_j) + \log \tilde{m}(f^*|\mathbb{M}_j)$$

where the first term of the right hand side is

$$\frac{(O - P_j)P_j}{2} \log 2 - \frac{\tilde{T}P_j}{2} \log \pi - \frac{P_j}{2} \log(\tilde{T}o_j + 1) - \frac{(\tilde{T} - P_j - O)}{2} \log |\psi_j| + \log \Gamma_{P_j} \frac{\tilde{T} + P_j - O}{2}$$

and the second term of the right hand side is

$$\frac{(O-P_j)P_j}{2}\log 2 - \frac{(O-P_j)(\tilde{T}-P_j)}{2}\log\pi - \frac{O-P_j}{2}\log(|W_j^*|) - \frac{(\tilde{T})}{2}\log|\psi_j| + \log\Gamma_{O-P_j}\frac{\tilde{T}}{2}$$

where

$$\begin{aligned}
\tilde{T} &= T - n_t \\
W_j^* &= \sum_{t=n_t+1}^{T} \tilde{f}_{j,t}\tilde{f}'_{j,t} \\
\psi_j &= \sum_{t=n_t+1}^{T} (\tilde{f}_{j,t} - \hat{\tilde{\alpha}}_j)(\tilde{f}_{j,t} - \hat{\tilde{\alpha}}_j)' + \frac{\tilde{T}}{\tilde{T}o_j + 1}(\hat{\tilde{\alpha}}_j - \tilde{\alpha}_j)(\hat{\tilde{\alpha}}_j - \tilde{\alpha}_j)' \\
\psi_j^* &= \sum_{t=n_t+1}^{T} (f^*_{j,t} - \hat{B}^*_{j,f}\tilde{f}_{j,t})(f^*_{j,t} - \hat{B}^*_{j,f}\tilde{f}_{j,t})'.
\end{aligned}$$

$\Gamma_d(.)$ denotes the d-dimensional multivariate Gamma function. All other variables are as previously defined. Hats indicate that parameters are the estimates obtained by linear regressions of (10) and (11).

In line with the recommendation by Chib et al. (2020) I use the model with the prior $\tilde{\alpha}_j|M_j \sim \mathcal{N}(\tilde{\alpha}_{j0}, o_j\Sigma_j)$ with

$$\tilde{\alpha}_{j0} = n_t^{-1}\frac{t=1}{n_t}\tilde{f}_{j,t},$$

where $n_t = tr \times T$ is the size of the training sample that I set to $tr = 10\%$ of the data. The model-specific o_j is calculated as

$$o_j = \frac{1-tr}{tr}L_j^{-1}sum(diag(V_{j0})/diag(\hat{\Sigma}_{j0})),$$

where V_{j0} is the negative inverse Hessian over $\tilde{\alpha}_j$ and $\hat{\Sigma}_{j0}$ the estimate of the covariance matrix Σ_j in the training sample.

References

Banerjee, A., & Davé, R. N. (2004). Validating clusters using the Hopkins statistic. *IEEE International Conference on Fuzzy Systems*, *1*, 149–153.

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

Chib, S., Zeng, X., & Zhao, L. (2020). On Comparing Asset Pricing Models. *The Journal of Finance*, *75*(1), 551–577.

Cross, G. R., & Jain, A. K. (1982). Measurement of Clustering Tendency. *IFAC Proceedings Volumes*, *15*(1), 315–320.

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

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

Fama, E. F., & French, K. R. (2017). International tests of a five-factor asset pricing model. *Journal of Financial Economics*, *123*(3), 441–463.

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

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A Test of the Efficiency of a Given Portfolio. *Econometrica*, *57*(5), 1121–1152.

Glorfeld, L. W. (1995). An Improvement on Horn's Parallel Analysis Methodology for Selecting the Correct Number of Factors to Retain. *Educational and Psychological Measurement.*, *55*(3), 377–393.

Hopkins, B. (1954). A new method for determining the type of distribution of plant individuals. *Annals of Botany*, *18*(2), 213–227.

Horn, J. L. (1965). A Rationale and Test for the Number of Factors in Factor Analysis. *Psychometrika*, *30*(2), 179–185.

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

Jensen, T. I., Kelly, B. T., & Pedersen, L. H. (2023). Is There a Replication Crisis in Finance? *The Journal of Finance*, *78*(5), 2465–2518.

Jolliffe, I. T. (1972). Discarding Variables in a Principal Component Analysis. I: Artificial Data. *Applied Statistics*, *21*(2), 160.

Jolliffe, I. T. (1973). Discarding Variables in a Principal Component Analysis. II: Real Data. *Applied Statistics*, *22*(1), 21.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141–151.

Kaufman, L., & Rousseeuw, P. J. (1990). Partitioning Around Medoids (Program PAM). *Finding groups in data: An introduction to cluster analysis* (pp. 68–125). Wiley & Sons Ltd.

Newey, W. K., & West, K. D. (1987). A Simple , Positive Semi-Definite , Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, *55*(3), 703–708.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.

Tang, T. M., & Allen, G. I. (2021). Integrated principal components analysis. *Journal of Machine Learning Research*, *22*, 1–71.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *63*(2), 411–423.