Scientific Portfolio
An EDHEC Venture

A Scientific Portfolio Publication

The Scientific Portfolio Risk Model

January 2023

The Scientific Portfolio risk model

Benoit Vaucher, PhD, CFA and Matteo Bagnara, PhD

2023 Q1

The EDHEC Scientific Portfolio (SP) platform assists institutional asset owners in understanding the risks of their portfolios. In this paper we present an implementation of the Instrumented Principal Component Analysis (IPCA) model [Kelly et al., 2019] for portfolios of equity portfolios, such as funds or mandates. We show that this model enable to decompose risk in many dimensions with high accuracy and stability. In contrast to current solutions, this model only requires widely available price data and extends easily to multi-region analysis. We present the model, address issues related to its calibration, and provide analytics that enable the decomposition of risks with a high level of historical and cross-sectional granularity. We also argue that the stability of this model makes its use possible not only in the context of risk analysis but also portfolio optimisation.

1 Introduction

The purpose of factor models is to identify and quantify how the value of a portfolio is exposed to the forces driving variations in asset prices. Their use in portfolio management is ubiquitous, from the identification of risks to their management via portfolio optimisation.

In this paper, we propose an application of the Instrumented Principal Component Analysis (IPCA) model [Kelly et al., 2018] to the risk analysis of portfolio of equity funds. As we are going to show, the IPCA model has a number of striking features that make it particularly fit for analyzing the risks of these instruments.

Equity funds tend to be invested across many industrial sectors and geographies, and their management involves different styles of active risk taking. Consequently, the risks involved in funds are subject to a potentially large number of factors.

For many investors who invest in portfolios of funds, the analysis of their risks is often problematic. Notably, the bottom-up approach that consists in using equity risk models to analyze the content of individual funds and then aggregate this information at the portfolio level, is difficult for a number of reasons.

First, detailed holding data at the fund level are often only precisely known to their managers. For asset owners relying on third

parties to manage parts of their portfolios, a detailed knowledge of the positions held in each fund at all time is often not available. Many commercial equity models that require detailed holding data are therefore difficult to exploit.

Second, the acquisition, maintenance and use of equity models is difficult. Expensive equity models need to be acquired for each region covered by the portfolio, and they are often not compatible between them, so that extra resources need to be expensed in order to combine their outputs. This creates a situation where the risk of each sub-portfolio is known, though the combined risks of the global portfolio are not well understood.

Against the backdrop, we have developed an implementation of the IPCA model that offers a solution to many of the key issues associated with the analysis of portfolios of funds. Importantly, it is designed to accommodate a great number of risk factors while also maintaining a high level of numerical stability.

As we are going to explain, this combination of depth and stability is one of the key innovation of this new type of models compared to their predecessors. This makes them very efficient not only for the purpose of risk analysis, but also, as we will illustrate, for portfolio optimisation.

The IPCA model uses as key input the financial characteristics of instruments. However, it has inherited the flexbility of previous cross-sectional models with respect to the definition of what constitute a characteristic. In contrast to equities, for which the original implementation of the model was intended, the financial characteristics of funds are not always, and one could even say, not often, readily available. We will show that using the instruments exposures to risk factors as characteristics, we are able to build very efficient risk models. This approach, that only require the instrument prices, enables the consistent estimation of financial characteristics across a wide spectrum of risk factors.

The consistent estimation of characteristics for a wide range of risk factors enables the construction of risk models for portfolios exposed to many different styles and geographies. The ability to examine the risk of a global portfolio with a single model sharply reduces the possibility of unintended risk exposures resulting from the aggregation of individual strategies.

Also, the availability of a risk model that easily applies to any fund or mandate opens up many possibilities. For instance, it also allows a consistent analysis of many investments across a large num-

ber of dimensions. This is particularly useful when analyzing large portfolios or searching for new investments, in which cases the comparability and consistency of risk information is key. Also, a detailed understanding of the contributions to the risk of the portfolio with a high level of granularity provides crucial information for decision making.

This paper is divided into two parts. In the first part, we present the IPCA model from a contextual point of view and explain how it compares and extend existing models. Then we explain how to estimate funds characteristics, and how to use them to calibrate the model. Although the main originality of this paper is to extend the application of the IPCA model to a new asset class, we propose a number of techniques that complement the original methodology. Notably, the calibration of the model in the context of funds present challenges that are not present for equities, for which we propose solutions. Also, we derive a number of original analytical formulas that enable the use of the model to decompose risk and returns into contributions that can be attributed to the different risk factors. We illustrate the use of these formulas on a universe of US mutual funds. We also provide discussions on the quantification of the model stability and why it lends itself well to portfolio optimisation tasks.

2 *Historical context*

Factor models have a long history in Finance. Their study has been an ongoing concern for financial economists since about the publication of the Capital Asset Pricing Model by Sharpe in 1964 [Sharpe, 1964]. The interest and motivation for the study of factor model arised from early observation that the variations in asset prices over a given period of time tend to share similarities. Stocks traded from the same market, for instance, tend to move in synchrony over any given day. It was also found that across the cross-section of returns, certain groups of stocks shared commonalities in the way their prices changed from one period to the other. These observations hinted the existence of forces, or factors, driving those variations. It is the intense interest of academics to understand the origin and nature of these drivers that led to the creation and development of factor models. The development of factor models was pioneered[1] by Sharpe in 1964 [Sharpe, 1964] and further developed notably by Ross [Ross, 1976] and Fama [Fama and French, 2015].

Although the first factor model by Sharpe only had one factor, the market itself, it was quickly extended to include multiple factors.

[1] For excellent introductions to factor models, see e.g. the Nobel lecture by Fama [Fama, 2013] or the guide to factor models edited by the CFA institute [Burmeister et al., 1994].

Factor models seek to explain the cross-section of asset returns by distinguishing the returns attributable to common forces to which the instrument is exposed, and the returns that are specific to the instrument itself. The forces driving the cross-sectional variations of prices are often refered to as systematic factors, as they influence all instruments. Once the impact of systematic factors has been accounted for, what is left of the returns is often refered to as specific.

In their most common form, factor models are linear and written in the form:

$$r_{i,t} = \boldsymbol{\beta}_{i,t} \boldsymbol{f}_t + \varepsilon_{i,t}$$

where $r_{i,t}$ is the return of the i^{th} instrument over a period ending at time t. The variable $\boldsymbol{\beta}_{i,t}$ is a vector of dimension $(1 \times K)$ that contains the exposures of the i^{th} instrument to the k^{th} systematic factors. Note that exposures are also referred to as *factor betas* or *factor loadings*. The returns of systematic factors between the times $t-1$ and t are denoted by the variable f_t, which is a vector of dimension $(K \times 1)$. Note that all variables are associated with individual instruments, while systematic factors are common to all instruments. Finally, the symbol $\varepsilon_{i,t}$ corresponds to the instrument's specific returns.

Because both factors and exposures are essentially unobservable, researchers have resorted to a wide range of assumptions and techniques to estimate them. Most of these approaches can be categorized into broadly two categories: statistical or fundamental.

Statistical models only require the asset prices to infer both exposures and factors. These models often use techniques inspired from principal component analysis to produce systematic factors (see e.g. [Stock and Watson, 2002]) and their associated loadings. Although statistical models tend to be precise, as measured by their R-squared, their factors do not carry a clear financial interpretation. This makes them well suited for tasks such as portfolio optimisation, which typically benefits from their precision[2], but inadequate for fundamentally driven financial analysis.

[2] Note that the covariance matrices obtained from statistical models are also often designed to be well-behaved when inverted, a must-have property for any model used in portfolio optimisation tasks.

Fundamental factors, on the other hand, posit that either factors or loadings are in fact observable. When it is the factors that are assumed to be observable, the model is refered to as *historical* or *time-series*. In time-series models, the factors returns often correspond to the returns of portfolios designed to represent the behaviour of a given ensemble of stocks. For instance, the returns of the Small-minus-Big (SMB) factors from Fama and French correspond to the returns of a portfolio that is long in stocks with small capitalization and shorts stocks of large companies (see e.g. [Fama and French, 1992] and reference therein). The loadings to the SMB factor cor-

respond to the regression coefficient of a multi-variate regression between the instrument's returns and the different factors. One of the known issues of time-series models is that they are limited to a few factors. This is a consequence of the presence of collinearities between these factors. If two factors are extremely correlated for instance, the regression will be indifferent to which factor it attributes the loading, so the attribution will be influenced by the specifics of the numerical procedure. In other terms, collinearities blur the causality link between the value of the loadings and the actual exposure of the instrument to the risk factor. Rather, the value of the loadings will depend on the constituents of the factor set. Because collinearities are difficult to avoid as the number of factors grows, this impairs the usability of time-series model for risk analysis when many factors are required.

Another type of risk model posits that the exposure are in fact observable, whereas the return series of the different factors must be inferred from the instruments exposures and their returns. This approach, which is well described e.g. by Grinold and Kahn in [Burmeister et al., 1994] or [Jacobs and Levy, 2021], is often referred as cross-sectional. In the context of cross-sectional models, the loadings are refered to as descriptors, or characteristics. Although more complex, this approach allows for the integration of many more factors, which is particularly well suited for the purpose of analyzing risk and returns from a large number of dimensions. Because of their flexibility and depth of analysis, cross-sectional models are widely used by practioners. They are commercialised by a number of providers, notably Barra.

The IPCA model pioneered by Fan and Liao [Fan et al., 2016] and Kelly and Pruitt [Kelly et al., 2019] mixes elements from statistical and fundamental cross-sectional models. As we will see later, its performance in terms of precision and stability shares many similarities with that of PCA-based models. But also, it also enables the decomposition of risks into contributions associated with fundamental characteristics. The IPCA model can thus be considered a bit of a hybrid. However, as we will see, its design solves many important issues associated with its predecessors and hence can be considered as an important innovation as far as factor models are concerned.

3 *Data*

In the sequel, we will illustrate the formulas and claims made in the different sections of this paper using the returns of the 400 largest

mutual funds present in the Morningstar database. We focus on a period starting from January 2000 to February 2022. The returns frequency is daily.

We will use fundamental and industrial factor returns over the same period, whose construction is described in section A.

4 *Model Specification*

The IPCA model specifies the returns $r_{i,t}$ of a given instrument from time $t-1$ to t as:

$$r_{i,t} = \boldsymbol{\beta}_{i,t-1}\boldsymbol{f}_t + \varepsilon_{i,t} \tag{1}$$

where $\boldsymbol{\beta}_{i,t} \in R^{1\times k}$ corresponds to the loadings at time t of the i^{th} instrument to the K risk factors $\boldsymbol{f}_t \in R^{K\times 1}$. The risk factors $\boldsymbol{f}_t$ will be refered to as statistical risk factors, a denomination that will be justified in the sections dedicated to the estimation of these factors. The symbol $\varepsilon_{i,t}$ denotes residual.

What makes the above definition different from traditional factor models is the way in which the definition of the loadings $\beta_{i,t}$ incorporate both fundamental and statistical information:

$$\boldsymbol{\beta}_{i,t} = \boldsymbol{z}_{i,t}\boldsymbol{\Gamma} + \eta_{i,t}$$

where $\boldsymbol{z_{i,t}} \in \mathbb{R}^{1\times L}$ is a collection of L fundamental characteristics associated with the i^{th} instrument. Characteristics are properties associated with each instrument that are expected to explain the cross-section of returns. They will be discussed in further detail in the next section. The matrix $\boldsymbol{\Gamma} \in \mathbb{R}^{L\times K}$ is one of the model's key parameters and finally, $\eta_{i,t}$ is the error term.

The loadings $\beta_{i,t}$ are therefore the result of a transformation of the fundamental characteristics by a multiplication with the matrix $\boldsymbol{\Gamma}$. The key role of the matrix $\boldsymbol{\Gamma}$ can be seen by condensing the model into a single line:

$$r_{i,t} = \boldsymbol{c}_{i,t}\boldsymbol{\Gamma}\,\boldsymbol{f}_t + \varepsilon^*_{i,t}$$

where the residual term becomes $\varepsilon^*_{i,t} = \eta_{i,t}\boldsymbol{f}_t + \varepsilon_{i,t}$. For ease of notation, we will drop the asterisk from the residual term. In what follows, the residual term is referred to as *specific return*, as corresponds to the part of the returns not related to systematic factors. The (delayed) vector of characteristics is defined as $\boldsymbol{c}_{i,t} = \boldsymbol{z}_{i,t-1}$.

The returns of a portfolio of N instruments are modeled as follows:

$$r_{w,t} = \boldsymbol{w}^T\boldsymbol{c}_t\boldsymbol{\Gamma}\boldsymbol{f}_t + \varepsilon_{w,t} \tag{2}$$

where $\boldsymbol{w}$ is the vector containing the weights of the different instruments and $\varepsilon^*_{w,t}$ corresponds to the specific returns of the portfolio. The dimension of the matrix $\boldsymbol{c}_t$ in this case is $(N \times L)$.

The matrix $\boldsymbol{\Gamma}$ is the key parameter of the model: **it corresponds to a mapping between the fundamental characteristics and the exposures to statistical risk factors**. It is so to speak the *Rosetta stone* that translates information from the fundamental world to the statistical world. The $\boldsymbol{\Gamma}$ matrix transforms information from the fundamental world to the statistical world. As discussed in [Kelly et al., 2019], the role of the gamma matrix is also to provide a noise reduction mechanism, as the elements of $\boldsymbol{\Gamma}$ associated with variables that do not contribute meaningfully to risk are effectively set to zero by the calibration procedure. The contribution of insignificant, noisy variables is thus effectively muted by the model[3].

[3] A more subtle mechanism also reported is the ability for the calibration procedure to use the lines of $\boldsymbol{\Gamma}$ to create linear combinations of possibly noisy characteristics in order to extract a signal while averaging out the noise.

Note that once the gamma matrix $\boldsymbol{\Gamma}$ and the factors $\boldsymbol{f}_t$ have been estimated, any instrument for which fundamental characteristics exist can be described by the model. This contrasts with e.g. PCA-based statistical models, for which an instrument needs to be part of the calibration set to be described by the model.

In the above specification, the characteristics are time-dependent, which is common with cross-sectional models. This contrasts with most historical models for which static characteristics are obtained from a set of pre-defined observable factors. It is useful to note that cross-sectional models can also be evaluated using static characteristics, which in many cases does not significantly impair their precision (see e.g. [Kelly et al., 2019] and [Fama and French, 2020]).

An important feature offered by the IPCA model, which is not present in cross-sectional models, is the possibility to set the number of statistical factors K to a value lower than to the number of fundamental characteristics L. As we will discuss in Sec. 9, reducing the number of factors is a key mechanism that allows a significant improvement of the model's stability. We have developed a criterion based on [Bai, 2003] and [Bai and Ng, 2009] to select the optimal number of factors.

5 *Perspective on cross-sectional factors and IPCA*

As we have mentioned in the introduction, the most common specification for factor models is the following linear relationship:

$$r_{i,t} = \boldsymbol{\beta}_{i,t}\boldsymbol{f}_t + \varepsilon_{i,t}$$

In the context of cross-sectional models, the loadings $\boldsymbol{\beta}_{i,t}$ are in fact observable characteristics[4], and the returns of the L latent risk factors $\boldsymbol{f}_t$ are obtained by solving

$$\boldsymbol{f}_t = \underbrace{(\boldsymbol{\beta}_t^T \boldsymbol{\beta}_t)^{-1}}_{\substack{\text{cross-sectional} \\ \text{covariance matrix}}} \underbrace{\boldsymbol{\beta}_t^T \boldsymbol{r}_t}_{\substack{\text{managed} \\ \text{portfolios}}} \tag{3}$$

where $\boldsymbol{\beta}_t$ is $(N \times L)$ matrix containing the L characteristics associated with the N instruments. The instruments returns at time t are denoted by the $(N \times 1)$ vector $\boldsymbol{r}_t$. Here the instruments can be said to be part of the calibration set used to estimate the model's factors. Note that the returns of cross-sectional risk factors are estimated at each time t.

[4] In the literature, the instruments properties that are used as loadings in cross-sectional models are sometimes called descriptors instead of characteristics.

As it can be seen from Eq. 3, the returns of the model's risk factors are formed by the multiplication of two components: the inverse of the cross-sectional covariance matrix of the characteristics and the returns of so-called managed portfolios. Cross-sectional covariance matrices measure the correlations between characteristics among instruments at a given time. For instance, if all instruments with a positive exposure to the "Value" risk factor also have a positive exposure to the "Small-Size" factor, then the cross-sectional correlation between "Value" and "Small-Size" will be positive. Managed portfolios are associated with each characteristic. They contain every instrument present in the calibration universe weighted by their exposure to the associated characteristics. Their returns at each time t thus corresponds to the exposure-weighted performance of all the instrument present in the calibration universe.

The construction of managed portfolios differs strongly from the typical construction of risk factors used in time-series models, such as e.g. the Fama-French five factors model, that typically include a limited selection of stocks with very high or very low exposure to the associated characteristics. In a sense the calculation of the managed portfolio returns use information from all stocks in the calibration universe, whereas the returns of historical factors depend on a small number of stock expected to be representative of the characteristics effects on performance.

The impact of the inverse of the cross-sectional covariance matrix on the factor returns is less straight forward. When the cross-sectional correlations between characteristics are very small or zero, the inverse of the covariance matrix merely rescale the returns of the managed portfolios. In this case, the factors returns correspond to those of the managed portfolios. As cross-sectional correlations between characteristics increase, the returns of the factors become

weighted combinations of managed portfolio returns. In this case, the relative contribution of each managed portfolio to the factors is proportional to the characteristics cross-sectional correlations.

Another important observation to be made from the above equation is that the existence of the factors f_t depends on the invertibility of the characteristics cross-sectional covariance matrix. Thus, when characteristics are cross-sectionally collinear, the factors cannot be evaluated. Because we evaluate characteristics from regressions, this could be the case if e.g. two fundamental risk factors are almost perfectly correlated. In a sense, when using characteristics obtained via regressions to time-series, the model stability also somewhat depends on the collinearities between fundamental risk factors, since the latter influence the cross-sectional correlations between characteristics. Correlated factors are likely to produce characteristics that are correlated cross-sectionally. However, we will see in the next sections that these stability issues are largely mitigated by the dimensionality reduction mechanism embedded in the IPCA model.

Another insight to be gained from formula 3 regards the definition of the $\mathbf{\Gamma}$ matrix. If we assume that characteristics are static and $\boldsymbol{\beta}_T = \boldsymbol{c}\mathbf{\Gamma}$, the equation 3 becomes:

$$\boldsymbol{r} = \underbrace{\boldsymbol{c}\mathbf{\Gamma}\boldsymbol{f}}_{\substack{\text{systematic} \\ \text{returns}}} + \boldsymbol{\varepsilon}$$

Here the variable $\boldsymbol{c}$ correspond to the matrix of static characteristics, $\mathbf{\Gamma}$ is $(L \times L)$, and $\boldsymbol{r}$ is $(N \times T)$. As before, we get

$$\mathbf{\Gamma}\boldsymbol{f} = (\boldsymbol{c}^T\boldsymbol{c})^{-1}\boldsymbol{c}^T\boldsymbol{r} \tag{4}$$

The gamma matrix $\mathbf{\Gamma}$ as well as orthogonalized factors $\boldsymbol{f}$ are easily obtained by performing a singular value decomposition on the right hand-side of Eq. 4:

$$\mathbf{\Gamma}\boldsymbol{f} \overset{SVD}{=} \underbrace{\boldsymbol{U}}_{(L\times L)}\ \underbrace{\mathbf{\Sigma}\boldsymbol{V}^{\boldsymbol{T}}}_{(L\times T)} \tag{5}$$

where we can identify $\mathbf{\Gamma} = \boldsymbol{U}$ and $\boldsymbol{f} = \mathbf{\Sigma}\boldsymbol{V}^{\boldsymbol{T}}$ which is orthogonal. Because $\boldsymbol{U}$ has the form of a rotation, the interpretation of the gamma matrix is indeed that it rotates the orthogonal factors back to the traditional cross-sectional factors.

Another insight that can be gained from the above formula is that the presence of the gamma matrix $\mathbf{\Gamma}$ actually offers the possibility to reduce the number of factors $\boldsymbol{f}_t$. Because the matrix $\mathbf{\Sigma}$ is diagonal, considering only the factors associated with the $K < L$ largest values of $\mathbf{\Sigma}$ generally yields a good approximation of the model's

parameters $\mathbf{\Gamma}$ and $\boldsymbol{f}$. This reduction in dimensionality does not affect definition of the systematic returns $\boldsymbol{c}\mathbf{\Gamma}\boldsymbol{f}$, despite $\mathbf{\Gamma}$ and $\boldsymbol{f}$ becoming $(L \times k)$ and $(k \times T)$, respectively. This reduction in dimensionality is only possible thanks to the presence of $\mathbf{\Gamma}$ right in between the characteristics and the (orthogonal) factors in the model specification. As we will explain in section 9, sparser models have better numerical properties, and therefore the ability to reduce the models dimensionality increases its stability.

This dimensionality reduction mechanism is actually quite a significant advance in the general context of cross-sectional models. One important criticism of these model is that their ability to accomodate many characteristics opens the door to the creation of overspecified models likely to suffer from numerical instability. Therefore, the ability to reduce the dimensionality of the model *without* reducing the number of characteristics solves this issue as it makes it possible to conserve a large number of dimensions for risk analysis while maintaining the model's numerical stability. A discussion on the relationship between the model's stability and the number of factors is given in Sec. 9.

Finally, note that the above approximation of gamma for static characteristics also constitutes an excellent starting point for the calibration procedure.

6 Characteristics

Characteristics, denoted by the letter $\boldsymbol{c}_t$, constitute the key ingredient of the IPCA model. The definition of a characteristics in the context of cross-sectional models is quite flexible: they are any instrument properties expected to explain variations in the cross-section of returns (see e.g. [Burmeister et al., 1994]). Alongside prices, they constitute the key inputs of the model.

As we are going to detail in the next sections, we use characteristics that correspond to the instrument's exposure to a set of risk factors. The characteristics of each instrument are thus obtained by regressing its returns against those of a set of risk factors. Well known risk factors are for instance the Fama-French five fundamental factors, or industry factors.

Historically, characteristics used in cross-sectional models fall broadly into two categories: fundamental and price-based characteristics. Our approach thus departs from the current practice as it only uses price-based characteristics, in the form of exposures, instead of a

mix of fundamental and price-based characteristics. As we are going to argue, there are several advantage to doing so.

Price-based characteristics are statistics estimated from the stock's price. The momentum or the market beta are two well-known characteristics estimated from the stock's price that are currently commonly used in cross-sectional risk models. The peculiarity of price-based characteristics in the context of cross-sectional risk models it that they generally depend on the price history. So although they are meant to describe the cross-section of returns at a time t, their value depend on the history of the price until that time. For instance, momentum of often evaluated over 9 months. Note that the market beta, a price-based characteristics often used in cross-sectional model, is also defined as an exposure to a risk factor. The use of price-based characteristics is thus quite common in cross-sectional model.

Fundamental characteristics are often based on financial ratios, such as the price-to-book ratio. Fundamental characteristics depend on the availability of accounting data and the frequency of earnings publications. Their comparability often requires re-adjustments depending on e.g. the industrial sector of the company or the accounting standards used to format the company's financial statement. Because the value of financial ratios does not vary between publications, their value at a time t correspond to the value of their latest publication, which can be up to a year old for most firms outside of the US.

The estimation of fundamental characteristics for portfolios of stocks is challenging for a number of reasons. First, up-to-date holdings information of often only available to the owner of the portfolio. Most funds do not report their holdings, and if they do, these reports are often outdated or incomplete, and do not take into account trades that may have significantly modified the fund's composition, and return patterns, between reports. This makes the estimation of financial ratios from the funds' underlying instruments difficult and prone to potentially large errors, both in terms of content and timeliness. It is also worth mentioning that the estimation of fundamental characteristics requires large and often expensive databases, which can be operationally problematic.

In contrast, measuring characteristics via regression analysis is consistent and unproblematic for most instrument in the investment universe[5]. Unless large portions of the funds historical prices are missing, funds characteristics obtained via regressions are generally available. Moreover, the large differences in amplitude between fundamental data points (e.g. company size in billions and prices-

[5] The availability of some financial ratios, on the contrary, is not always guaranteed due to the different reporting practices in different sectors

to-book in fractions) that normally require normalisation procedures to be used the model's calibration procedure are not present when dealing with exposures.

The timeliness of funds characteristics is also unproblematic. Contrary to financial ratios whose publication is quarterly, characteristics derived from prices include the latest market information.

Finally, the estimation of exposures from regression yields a confidence interval around the value of the characteristics. As we are going to detail later, confidence intervals are straight-forward to obtain and provide information about the robustness of the estimate going forward, an information that is often absent with financial ratios, although uncertainty around their values does also exist.

7 Risk Factors

The fundamental risk factors that we are going to use to evaluate funds characteristics include of course the market itself, but also a number of style and industry risk factors. The methodology used to evaluate characteristics will be detailed in the next section.

Generally speaking, risk factors correspond to the returns of portfolios of stocks that are representative of the behaviour of a certain group of stocks. A well-known set of factors is the Fama-French five factors model for instance. There are many types of risk factors that are used to analyse the returns of portfolios. Here we consider so-called *style* factors and *industry* factors.

The style factors we use are referred to as *rewarded* factors, because of the well-established relationship with long-term excess returns. They include the Value, Small-Size, Low-Volatility, Profitability, Momentum and Investment factors (see [Amenc and Goltz, 2016] or [Freyberger et al., 2020] for excellent reviews). Their construction is detailed in the Appendix A.

Industry factors are referred to as *unrewarded*. Unrewarded risk factors explain risk but they are not expected to generate excess returns in the long-term. The most important unrewarded risk factors in the context of equities are industrial sectors. Exposures to industrial sector do indeed explain a large amount of commonalities in the cross-section of stocks returns. However, contrary to rewarded factors, they do not confer a performance advantage in the long term. The construction of industry factors is detailed in the Appendix A.

Because we are dealing with portfolios of equities, the market

factor is de-facto the prominent source of risk. In this paper, all fundamental factors are designed to be orthogonal to the market factor. In this ways, the market risk of each instrument is captured uniquely by its market characteristic. The characteristics of instruments to non-market risk factors are thus associated to their active risk.

8 Estimation of Characteristics

The characteristics $(\boldsymbol{c}_t)_{i\ell}$ of the i^{th} instrument associated with the ℓ^{th} factor are obtained at each time t by solving the following linear model

$$\boldsymbol{r}_{i,t} = \sum_{\ell} (\boldsymbol{c}_t)_{i\ell} \; \boldsymbol{F}_{\ell,t} + \boldsymbol{\varepsilon}_{t,i}$$

As we have mentioned before, the factor we use are orthogonal to the market factor, and therefore here $\boldsymbol{r}_{i,t}$ denotes to the instrument active returns over a period of fixed size ending at time t and $\boldsymbol{F}_{\ell,t}$ the returns of the ℓ^{th} risk factor over the same period. The active returns correspond to the residual returns of a regression between the instrument (absolute) returns $R_{i,t}$ and the market factor, that is:

$$\boldsymbol{R}_{i,t} = (\boldsymbol{c}_t)_{im} \; \boldsymbol{F}_{m,t} + \boldsymbol{r}_{i,t}$$

with $\boldsymbol{F}_m$ the market factor.

We estimate the parameters via a weighted least-squares regression (WLS) for which an analytical solution exists:

$$(\boldsymbol{c}_t)_{i\ell} = (\boldsymbol{F}_\ell \boldsymbol{W} \boldsymbol{F}_\ell^T)^{-1} \boldsymbol{F}_\ell^T \boldsymbol{W} \boldsymbol{r}_i$$

The matrix $\boldsymbol{W}$ is diagonal and contains the weight of each day in the regression. We use exponentially decaying weights $(\boldsymbol{W})_{tt} = \lambda^t$ with $\lambda = \exp\{-\frac{\ln 2}{\tau}\}$ so that the more recent period has a relatively larger importance in the evaluation of characteristics. More importantly, this weighting scheme mitigates the so called "window" effect when a particularly volatile period exits the evaluation window. With decaying weights, the impact of past events is progressively scaled down as time passes. In the sequel, we use a half-life τ of approximately 18 months.

Note that there should be no need to evaluate the market beta separately since all factors are orthogonal to the market factor. However, although the risk factors are designed to be orthogonal, some residual correlation with the market is not entirely unavoidable. Estimating the market exposure first and using the residual returns, that are orthogonal to the market factor, to evaluate the other characteristics ensures that market risk is not transfered to other factors. This

is important given the relatively large portion of the total risk that is attributed to this factor.

As we mentioned before, characteristics obtained from regressions possess a confidence interval that stems from the coefficient standard error, which is readily available given by

$$(\mathrm{SE}_t)_{i,\ell} = \sqrt{\frac{\sum_t \varepsilon_{t,i}^2}{(T-2)\sum_t F_{\ell,t}^2}}$$

The average value and standard error of the characteristics belonging to the 400 largest US mutual funds are shown in Fig. 1. The length of the window chosen for this calculation is 5 years ending in February 2022.

For multi-region portfolios, the characteristics of instruments belonging to different regions are obtained using their local factors. Exposures to the other regions' risk factors are set to zero. The matrix of characteristics for these such case will look like the table below:

	US Factor 1	US Factor 2	EU Factor 1	EU Factor 2
US Funds 1	0.10	0.5	**0**	**0**
US Funds 2	0.4	-0.2	**0**	**0**
EU Funds 1	**0**	**0**	0.3	-0.76
EU Funds 2	**0**	**0**	-0.1	0.5

Table 1: Representation of a matrix of characteristics for two US and EU instruments exposed to two regional risk factors. methodology

In this fashion, the characteristics of all instruments are evaluated using their local risk factors. Regional models can thus be built using prices and matrices of characteristics such as the one above where both the instrument prices and local risk factors are translated to the investor's currency before evaluating the characteristics.

Research considerations on the estimation of characteristics

Fundamental ratios versus covariances Apart from historical practice, the use of exposures in cross-sectional models is not forbidden by principle, quite on the contrary. In fact, many characteristics used in commercially available cross-sectional models are obtained from historical data, or are actually exposures: market beta, momentum, fundamental ratios based on trailing prices, etc.

In practice, exposures and fundamental ratios are used in a large number of application related to risk management, such as hedging

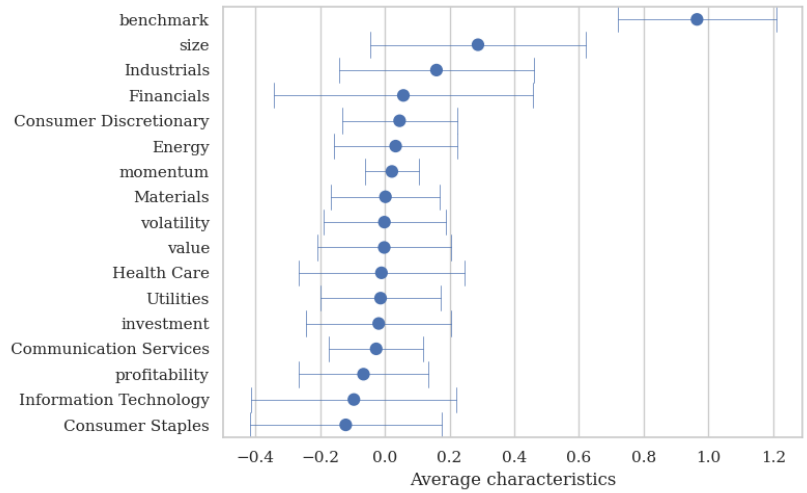

Figure 1: Distribution of the characteristics among the 400 largest US funds for a selection of risk factors.

or portfolio management. None seems to be provide systematically superior results, which indicates that both types of characteristics provide useful information about risk. As we are going to illustrate in 10, the use of exposures as characteristics leads to cross-sectional models with high precision levels, almost on par with price-based statistical PCA models.

The academic literature does not either offer definitive conclusions as whether betas or fundamental loadings constitute better risk characteristics. The difference between loadings and financial ratios has been discussed primarily in the context of asset pricing. The determination of what constitutes a characteristic is important as it boils down to what constitutes the essence of equity premia: are premia the result of an association with a group of stocks sharing a given price pattern, i.e. the result of covariances, or the consequence of an intrinsic stock characteristics unrelated to return patterns. In other terms, is it right to call a stock "Value" because its price behaves similarly to other Value stocks, or because it has a low price-to-book ratio.

In a famous series of papers [Daniel and Titman, 1998, 2005], Daniel and Titmann remarked in that some stocks with a low price-to-book did not actually covary with the Value factor *although their Value premium was still present*. To draw this conclusion, the authors estimated the loadings of all stocks to Fama's HML factor and compared them to the rank of their price-to-book (P/B). They found that some stocks with a high B/P but a low loadings still had a premia, thus suggesting that financial ratios determined the association with

premia.

This finding was commented by Davis and Fama [Davis et al., 2000] who showed that first of all, stocks that have exposures and fundamental ratios in opposite sides of the distribution are rare. With the exception of the most recent period, over which the study of Titman was performed, covariances with risk factors is actually what statistically determines its expected returns, not fundamental ratios. So empirically, what emerges from these studies is that neither covariances nor fundamental ratios seems to better determine the behaviour of a stock at all time.

Stability of characteristics compared to fundamental ratios An necessary property of characteristics is that they need to be somewhat stable over time[Menchero and Wang, 2011]. In this context, stability means that the characteristics do not exhibit radical changes over short period of time. Indeed, characteristics that change fast have little appeal from a risk and portfolio management perspective, as it may require frequent holdings adjustments to control it. One way to measure the stability of characteristics is to measure their cross-sectional correlation from e.g. one month to the next:

$$\rho_{\ell t} = \frac{\sum_i ((\boldsymbol{c}_t)_{i\ell} - (\bar{\boldsymbol{c}}_t)_\ell)((\boldsymbol{c}_{t+1})_{i\ell} - (\bar{\boldsymbol{c}}_{t+1})_\ell)}{\sqrt{\sum_i ((\boldsymbol{c}_t)_{i\ell} - (\bar{\boldsymbol{c}}_t)_\ell)^2}\sqrt{\sum_i ((\boldsymbol{c}_{t+1})_{i\ell} - (\bar{\boldsymbol{c}}_{t+1})_\ell)^2}}$$

with $(\bar{\boldsymbol{c}}_t)_\ell$ the average of the ℓ^{th} characteristic at time t. It is generally considered that an average stability of less than 0.8 over time is too unstable for model inclusion, while values of 0.9 and above are considered desirable [Menchero and Wang, 2011].

We found that the average stability of the characteristics estimated from risk factors is on average 0.98 month-over-month over ten years, which makes them perfectly fit for use in a model from a stability viewpoint. We have estimated this average by measuring the characteristics of the largest 1200 stocks of the US market over a period of 15 years. The same level of stability (0.95) is observed when a time horizon of a quarter is used.

For the sake of comparison, we have measured the stability of the price-to-book (P/B) for the same sample of stocks over the same period. We have normalized the ratios by applying a cross-sectional z-score on their values, as it is done customarily in equity cross-sectional models. This normalisation is important, as otherwise the stability of the ratio would be severely impacted by the unstability of the price itself. After normalisation, we found that the average stability of the (P/B) ratio from month-to-month is 0.9, a value somewhat lower than its price-based counterpart. We have also estimated

the stability of normalized prices, which is 0.93. This value provides an upper bound to the stability of price-based ratios. Effectively, the stability of the price itself corresponds that of a fundamental whose denominator is fixed.

This analysis shows that for characteristics based on price ratios, the stability of characteristics estimated from risk factors is likely to be somewhat higher than their fundamental counterparts. The relative stability of other characteristics depends on their very definition. For instance, the market capitalisation of a firm may be very unstable whether it is estimated using the latest market value, or only once at the beginning of the year. All in all, stability of characteristics between the fundamental and price-based approach is broadly similar and does not favour one approach over the other.

Multi-variate versus univariate regressions Notice that all characteristics need not be estimated at once in a single regression. It is possible to obtain characteristics by performing the regression defined in Eq. 8 for different set of factors (e.g. style, industries, region) or even for each factor separately. This possibility is important because it enables the estimation of characteristics over small groups of risk factors.

Performing multi-variate instead of univariate regressions, e.g. when estimating characteristics using one risk factor at a time, has a number of numerical advantages. First, it reduces the standard error of the estimates. Multi-variate regressions involving either fundamental factors, e.g. Value, Size, Momentum, or industry factors tend to produce smaller residues.

Second, we found that in some cases, univariate regressions were cross-sectionally very correlated. As we have explained before, this is problematic as the cross-sectional covariances matrix between characteristics needs to be inverted in order to obtain the models factors (see Sec. 5). This numerical operation is sensitive to the matrix stability, or in other terms, its distance to singularity. In general, large cross-sectional correlations between characteristics tend to impair the stability of the matrix

To illustrate the impact of the estimation method on the cross-sectional correlation matrix of characteristics, we select 400 US funds and measure their exposures to a set of six fundamental risk factors over five years, between the end of March 2017 and the end of March 2022 using two methods. The first method estimates the model characteristics by regressing the active returns on each risk factor individually. The corresponding cross-sectional covariance matrix of the regression coefficients is displayed in the table below.

	volatility	IT	value	momentum
volatility	1.00	0.33	-0.57	0.45
IT	0.33	1.00	-0.92	0.94
value	-0.57	-0.92	1.00	-0.96
momentum	0.45	0.94	-0.96	1.00

Table 2: Cross-correlations between characteristics estimated using a univariate regression.

The second method is a multi-variate regression. The cross-sectional covariance matrix associated with the characteristics obtained using this method is displayed in table 3.

	volatility	IT	value	momentum
volatility	1.00	-0.47	-0.17	-0.39
IT	-0.47	1.00	-0.05	0.35
value	-0.17	-0.05	1.00	0.05
momentum	-0.39	0.35	0.05	1.00

Table 3: Correlations between characteristics measured with multivariate regression

When two sets of characteristics are highly correlated cross-sectionally, the cross-sectional covariance matrix is pushed towards singularity, which makes it numerically unstable. This instability has the potential to impair the estimation of the model's factors. We found that by grouping risk factors when estimating the characteristics, the properties of the characteristics were improved. Compared to characteristics obtained with univariate regressions, their cross-sectional correlation matrix is much better behaved and more stable under inversion, which should benefit the stability of the model's factors estimation.

Static versus Dynamic Characteristics In the model specification, instrument characteristics are dynamic, in the sense that they can have different values at each time t. In the context of equities, price-based characteristics change every period, whereas characteristics based on fundamental ratio remain the same between release.

When using regression analysis to obtain characteristics, the length of the historical track-record may limit the availability of characteristics. In this case, it is possible to set the value of characteristics to a fixed value at each time t. This value may correspond to the regression coefficient over the whole period of interest, or the average of the availablel dynamical characteristics.

Note that models calibrated with dynamic characteristics cannot

be used on instruments with only static characteristics. Separate risk models need to be calibrated for either for static or dynamic characteristics. We have found that in practice, for periods of time of up to five years, static and dynamic models deliver equally good precision.

9 *Estimation of the models parameters*

The IPCA model contains two parameters that must be estimated: the Gamma matrix, $\mathbf{\Gamma}$ and the statistical risk factors, $\boldsymbol{f}_t$. The number of factors K must be set before the calibration procedure.

Any calibration method seeks parameters that improve the model's sample fit. That is to say, the parameters $\mathbf{\Gamma}$ and $\boldsymbol{f}_t$ must reduce the difference between the model's systematic returns and the actual returns of each instrument present in a given calibration set. Mathematically, this objective is expressed into the following optimization program:

$$\begin{aligned}(\mathbf{\Gamma}^\star, \boldsymbol{f}_t^\star) &= \underset{\mathbf{\Gamma}, f_t}{\operatorname{argmin}} \sum_{i,t} (r_{i,t} - \boldsymbol{c}_{i,t}\mathbf{\Gamma}\boldsymbol{f}_t)^2 \\ &= \underset{\mathbf{\Gamma}, f_t}{\operatorname{argmin}} \sum_{t} (\boldsymbol{r}_t - \boldsymbol{c}_t\mathbf{\Gamma}\boldsymbol{f}_t)^T (\boldsymbol{r}_t - \boldsymbol{c}_t\mathbf{\Gamma}\boldsymbol{f}_t)\end{aligned}$$

where $\boldsymbol{c}_t \in R^{N\times L}$ and $\boldsymbol{r}_t \in \mathbb{R}^{N\times 1}$.

In order to obtain the optimal $\mathbf{\Gamma}$ and f_t, the authors of the IPCA model [Kelly et al., 2019] have put forward the use of an alternating least-squares method. As its name suggests, this method consists of alternating between the estimation of the factors while fixing the values of gamma matrix, and the estimation of the gamma matrix using fixed factors. The rate of convergence, the stability and asymptotic properties of this method are studied quite extensively by Kelly et al. [Kelly et al., 2019] and compare very favorably to competitive models such as the PCA model. For excellent analyses of the model's asymptotic properties, see also [Bai et al., 2016] and [Bai, 2003].

Although somewhat technical, the purpose of the current section is not only to detail the calibration procedure of the parameters, which broadly follows the presentation made in [Kelly et al., 2019], but also to illustrate two of the model's less visible advantages. First, its ability to be estimated on unbalanced data panels. Second, the possibility to obtain statistical factors that are orthogonal. Both of these properties stem from the model estimation methodology.

Finally, note that the N instruments used to calibrate the model are said to be part of a so-called calibration set. Because the application of the IPCA model to funds raises specific issues with respect to the selection of the calibration set, this topic will be discussed in detail in Sec. 9.

Alternating least squares

The alternating least-squares algorithm begins with an initial guess for the variables $\boldsymbol{f}_t$ and $\boldsymbol{\Gamma}$, and then iterates between updates of $\boldsymbol{\Gamma}$ and $\boldsymbol{f}_t$ while holding the other fixed. After each update, a transformation R is applied to $\boldsymbol{\Gamma}$ and $\boldsymbol{f}_t$. The nature and definition of this transformation will be detailed in Sec. 9. When the difference between the previous and new variables becomes smaller than a predefined threshold, the algorithm terminates.

In our implementation, the initial guess for the variables is simply the analytical solution of a static version of this model as defined in Fan [Fan et al., 2016]. Another excellent initial guess is presented in Sec. 5.

Algorithm details The first step of the algorithm consists in freezing the value of the previous $\boldsymbol{\Gamma}$ while estimating the values of $\boldsymbol{f}_t$ at each time t. This step consists in finding a solution to

$$\boldsymbol{f}_t^{new}(\boldsymbol{\Gamma}_0) = \underset{\boldsymbol{f}_t}{\operatorname{argmin}} (\boldsymbol{r}_t - \boldsymbol{c}_t \boldsymbol{\Gamma}_0 \boldsymbol{f}_t)^T (\boldsymbol{r}_t - \boldsymbol{c}_t \boldsymbol{\Gamma}_0 \boldsymbol{f}_t) \tag{6}$$

which is readily available[6] as

$$\boldsymbol{f}_t^{new} = (\boldsymbol{\Gamma}_0^T \boldsymbol{c}_t^T \boldsymbol{c}_t \boldsymbol{\Gamma}_0)^{-1} \boldsymbol{\Gamma}_0^T \boldsymbol{c}_t^T \boldsymbol{r}_t \tag{7}$$

[6] The optimisation program of Eq. 6 is identical to the one used to solve the simple regression $\boldsymbol{r}_t = \boldsymbol{c}_t \boldsymbol{\Gamma}_0 \boldsymbol{f}_t + \varepsilon_t$

Here the number of points present in this regression is equal to the number of instruments. The precision of $\boldsymbol{f}_t^{new}$ is thus asymptotically proportional to $\sqrt{N}$. Because this regression is applied cross-sectionally at each time point, the number of instruments with valid data at each time can vary.

The next step of the estimation algorithm consists of keeping $\boldsymbol{f}_t^{new}$ constant while estimating $\boldsymbol{\Gamma}_{new}$ using the following formula:

$$\operatorname{vec}(\boldsymbol{\Gamma}) = \left[\sum_{i,t} (\boldsymbol{f}_t^T \boldsymbol{f}_t \otimes \boldsymbol{c}_{i,t} \boldsymbol{c}_{i,t}^T) \right]^{-1} \sum_{i,t} (\boldsymbol{f}_t \otimes \boldsymbol{c}_{i,t}^T) \operatorname{vec}(r_{i,t}) \tag{8}$$

A derivation of the above formula is provided in Appendix B.

As we can now see, the update of both variables is *resilient to the presence of missing values*, as only the characteristics of instruments with available data at each time t are used while instruments with missing data are simply left out. The algorithm is thus accomodating so-called *unbalanced* data panels.

Data panels are called "unbalanced" when returns and characteristics are not available at every point in time and for each instrument. This can be seen from the above definitions of Γ as a sum of characteristics multiplying factor returns. The sum takes only data that is available, and there is no requirement for each instrument to have data available at each point in time. The same remark applies to the estimation of the factors returns.

Many real-life data panels often have missing data points, so the ability to use unbalanced data panels to estimate a model's parameters is important in practice. Because of this flexibility, it is not necessary for all instruments in the calibration set to have data available at each point in time. This dramatically increases the data available to fit the model. This is in stark contrast to PCA-based statistical models for which missing data must be filled in by, for example, replacing missing returns with the sector average or zero. The methods used to replace missing data can have a strong impact on the parameters and affect the model's precision.

Another less visible advantage of using unbalanced data panels is that the resulting models are less affected by survivor bias. When instruments are required to cover the full period, the risks associated with instruments that have disappeared over the period are not taken into account when estimating the model's parameters.

After determining the values of $\mathbf{\Gamma}$ and $\boldsymbol{f}_t$, the final step is to apply a transformation R of dimensions $(K \times K)$ on both parameters. This reason for this last step and the definition of the transformation are detailed in the next section.

Rotational invariance and factors orthogonality

From its specification, one can see that the model is invariant under the transformation

$$\begin{aligned} \mathbf{\Gamma}' &= \mathbf{\Gamma}\, R \\ \boldsymbol{f}'_t &= R^{-1} \boldsymbol{f}_t \end{aligned}$$

where R is a $k \times k$ matrix. Because

$$\mathbf{\Gamma}' \boldsymbol{f}'_t = \mathbf{\Gamma} \boldsymbol{f}_t,$$

the above transformation leaves the model unchanged and reflects the so-called *rotational invariance* of the model parameters. Instead of being uniquely defined, the parameters, $\mathbf{\Gamma}$ and $\boldsymbol{f}_t$, are defined *up to a transformation*. This transformation changes the *relative* values of the parameters, but leaves the model unchanged. That is to say, optimal parameters, which maximize the sample fit, remain optimal after a rotation. The above rotation thus does not have any impact on the model precision. It simply modifies the relative definition of the parameters.

A concept of interest in this context is the existence of so-called *uniquely identifiable parameters*. Because parameters can always be defined up to a rotation, it is possible to use these rotations to make parameters have certain properties. For instance, it is possible to use the transformation R to make the factors $\boldsymbol{f}_t$ orthogonal[7].

[7] The SVD decomposition of $\boldsymbol{f}_t = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ implies that setting $\boldsymbol{R} = \boldsymbol{U}$ produces factors $\boldsymbol{f}'_t$ that are orthogonal

What should be noted here is that using R to impose certain properties on the parameters constrains its definition. For instance, imposing the orthogonality of the factors f_t implies $K(K-1)/2$ constraints, namely that the multiplication of each columns $i \neq j$ must be zero. When the number of conditions placed on R becomes equal to its number of components K^2, it becomes uniquely defined, which actually lifts the rotational indeterminacy. Another important consequence of this is that when R is unique, then the parameters themselves also become unique. In this case, they are said to be *uniquely identifiable* or *rotationally invariant*. A simple way to obtain uniquely identifiable transformations is thus to find a number of constraints on the parameters that is equal to the number of elements present in the transformation matrix R.

In the context of the present optimisation procedure, uniquely identifiable parameters have a great property: their convergence to optimality is uniform and guaranteed. It is therefore desirable to maintain the parameters uniquely defined during optimisation procedure.

However, the way calibration procedure work is that at each step, a small transformation is applied on the parameters so as to make them closer to their optimal value. If the initial parameters are part of a class of uniquely identifiable parameters, this small transformation will make them leave their class, and consequently impair the convergence properties of the algorithm. This is why the parameters must be brought back to their uniquely identifiable class after each step.

One of the uniquely identifiable class of parameters proposed in [Kelly et al., 2018] consists of imposing to to f_t to be orthogonal and

$\mathbf{\Gamma}$ to be orthonormal, that is:

$$\mathbf{\Gamma}'\mathbf{\Gamma}'^{T} = 1$$
$$f'_t f'^{T}_t = \boldsymbol{D}$$

where $\boldsymbol{D}$ is a diagonal matrix. The transformation R that transports the variables $\mathbf{\Gamma}$ and f_t to their rotationally invariant class is defined as follows:

$$\mathbf{\Gamma} \overset{Cholesky}{=} \boldsymbol{A}^T \boldsymbol{A}$$
$$\boldsymbol{A}\, f_t f_t^T\, \boldsymbol{A}^T \overset{SVD}{=} \boldsymbol{B}\boldsymbol{D}\boldsymbol{B}^T$$
$$R = \boldsymbol{A}^{-1}\boldsymbol{B}$$

One can easily check that:

$$\begin{aligned}
\mathbf{\Gamma}'\mathbf{\Gamma}'^{T} &= \mathbf{\Gamma} R (\mathbf{\Gamma} R)^T = \boldsymbol{B}\boldsymbol{B}^T = \mathbb{1} \\
f'_t f'^{T}_t &= R^{-1} f_t (R^{-1} f_t)^T \\
&= \boldsymbol{A}\boldsymbol{B}^{-1} f_t (\boldsymbol{A}\boldsymbol{B}^{-1} f_t)^T \\
&= \boldsymbol{D}
\end{aligned}$$

The proof that this transformation is indeed unique is provided in [Kelly et al., 2019]. As simple way to understand why these conditions produce a rotationally invariant class is to count the number of constraints. As we already mentioned, he diagonality condition $f_t f_t^T = \boldsymbol{D}$ involves $k(k-1)/2$ constraints. The orthonormality constraint is a combination of a diagonality constraint with the addition of k constraints as the multiplication of identical column must equal one. It thus totals a number $k(k-1)/2 + n$ of constraints. The sum of constraints associated with these two conditions thus leads to a total of

$$k(k-1)/2 + k(k-1)/2 + n = k^2$$

constraints. This number is sufficient to constrain every element of the R transformation, hence leading R to be part of a uniquely identifiable class.

Beyond their convergence guarantees, there are other advantages that come with imposing certain properties to the model's parameters. As shown in [Kelly et al., 2019], certain uniquely identifiable classes enable to fix certain elements of the gamma matrix so as to guarantee a pre-defined financial meaning to the factors f_t. In Appendix E, we propose a new rotationally invariant class of parameters that actually reduces the variations between the optimal gamma and a reference gamma. This is useful for instance to study the impact of data outliers on the stability of the gamma matrix, or make risk decompositions more consistent across periods.

Finally, note that imposing orthogonality to the factors f_t simplifies in a number of analytical calculations related to the application of the model. Notably, in the context of robust portfolio allocation, factors orthogonality dramatically reduces the complexity of many optimisation programs (see e.g. [Goldfarb and Iyengar, 2003]). The use of the IPCA in the context of robust optimisation goes beyond the intent of this paper and will be developed elsewhere.

Number of statistical factors

The number of implicit factors $K \leq L$ is a technical parameter of the model that has to be set before the calibration procedure. As we are going to discuss now, the number of factors has a strong impact on the model's stability, and also its informational content.

The stability of a factor model is linked to the stability of its associated covariance matrix. This important topic has been studied quite extensively, see e.g. [Ledoit and Wolf, 2003] and references therein.

Generally speaking, the stability of a matrix is related to the amplitude of its smallest eigenvalues[8]. In the context of covariance matrices, the presence of numerically small eigenvalues often reveal a high sensitivity to small changes of the model's parameters, a well-known signature of instability. When some their eigenvalues become vanishingly small, which happens when the returns of instruments are highly collinear or the number of instruments becomes too large compared to the number of time points, the covariance matrix may even becomes singular. Singularity drastically impairs the useability of the model for both for risk analysis and portfolio optimisation tasks.

[8] The qualification of a matrix stability is a theme that has been heavily studied in mathematics and is often referred to as the "conditioning" of a matrix. See e.g. "Numerical Linear Algebra" by Trefethen and Bau (SIAM, 1997).

The reason why reducing the number of factors stabilizes the covariance matrix is that it effectively eliminates small eigenvalues. This effect will be discussed more in depth in Sec. 11 where we present a widely used criterion that quantifies the stability of a matrix and show how reducing the number of factors contributes to its improvement.

Away from mathematical considerations, the benefits of reducing the number of factors can be understood from an informational point of view. Model calibration procedures prioritizes the information found in the calibration set by order of importance. Patterns explaining the large price variations will be considered more important than patterns explaining small price variations. In order to store the information it acquires on the instrument returns, the calibration procedure will use the model's parameters. So to speak, the model's

parameters constitute the hard-drive of the model, its informational capacity.

If a model only contains a few factors, the calibration procedure will use the limited informational capacity of the model to store only the most important, most stable return patterns. By allowing less factors into a model, the calibration procedure is so-to-speak forced to focused on significant patterns because it does not have the informational capacity to store less significant patterns, that are also more likely to be actually be associated with noise. When more factors are let in, the model thus runs the risk of becoming *over-fit*, by which we mean that the model is eventually able to store all return patterns, which includes sample noise.

In our case, reducing the number of implicit factors, and thus the informational capacity of the model, to a number $K \leq L$ reflects the fact that some characteristics may actually convey the same information and thus do not need extra capacity to be stored. When characteristics actually exhibit strong cross-sectional correlations, the information they convey about returns can be compressed in a fewer number factors.

The problem at hand when selecting the number of factors is to find the right tradeoff between reducing the number of factors and reducing it too much. Reducing the number of factors invariably reduces the precision of the model in and out-of-sample, which is also problematic. Defining the right number of factors thus essentially amounts to allowing parameters as long as they meaningfully contribute to the model precision. If the model has captured the most important return patterns, adding more parameters, and thus allowing for less important patterns to be stored, should not contribute to increase precision. This is illustrated in Fig. 2, where we have plotted the total R-Squared of risk models with different values of K.

When the plateau is reached, the addition of factors only marginally improves the precision and the model becomes prone to over-fitting.

The typical approach discussed in the literature to evaluate the right number of factors is to define an information criterion that balances the informational capacity of a model with its precision. Information criteria are thus often composed of one part that increases with the model's accuracy, and another part that decreases with the number of parameters. One of such criteria, that was proposed by [Bai, 2003] and [Bai and Ng, 2009] in the context of static factor models, is defined as follows :

$$IC(k) = \log S(k) + k\, g(N, T)$$

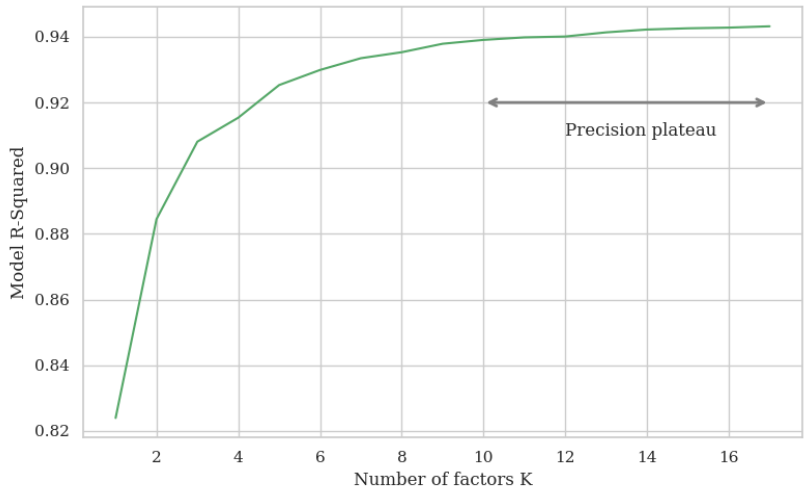

Figure 2: Model's precision reaches a plateau past a certain number of factors. The models for different values of the parameter K are calibrated on the 400 largest US mutual funds.

with

$$S(k) = (NT)^{-1} \sum_{i,t} \varepsilon_{i,t}^2(k)$$

$$g(N,T) = \frac{N+T}{NT} \log\left(\frac{NT}{N+T}\right)$$

where $S(k)$ is the sum of the model's squared residuals $\varepsilon_{i,t}$ for all instruments i when the model contains k factors. It is a measure of the model's goodness of fit. The term $g(N,T)$ is a penalty that increases with the number of factors.

Although alternative definitions of the S and $g(N,T)$ functions are possible, one important condition is that $g(N,T) \to 0$ as $N,T \to \infty$. This type of information criterion is used quite extensively in the literature on factor models (see e.g. [Amengual and Watson, 2007]).

One important point is that the definition of the information criteria used to select the number of factors must reflect the purpose of the model. Models for which precision is important will require information criteria where the relative balance between the functions S and g gives more importance to the model's goodness of fit. Models for which informational efficiency is important will increase the relative weight of g.

In our case, we have slightly adjusted the criterion proposed in [Bai and Ng, 2009] to favor somewhat sparser models. Instead of the above $g(N,T)$ function, we use the following scaled-up version:

$$IC_{ESA}(k) = \log S(k) + 4\,k\,g(N,T)$$

Note that multiplying the function $g(N,T)$ by a constant $\alpha \sim O(1)$

does not modify its convergence properties - it merely puts more weight on the part of the equation that favors sparsity. As illustrated in Fig. 3, this change has the effect of moving the optimal point of the information criterion to the left, which favours models with less factors.

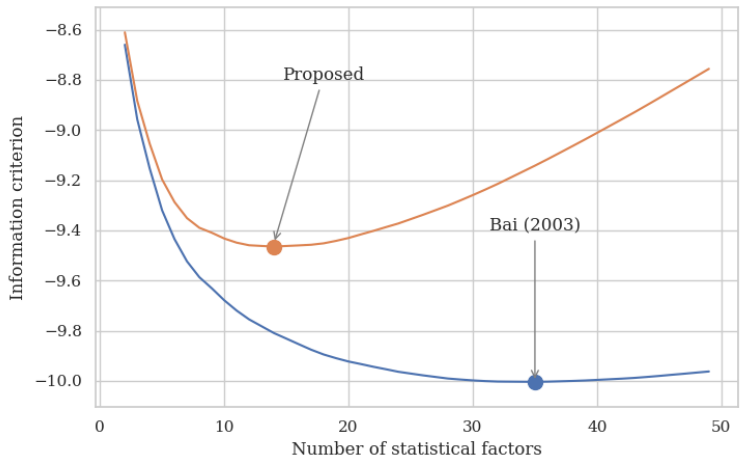

Figure 3: Comparison between the information criterion from [Bai and Ng, 2009] and the present proposal. The dots locate the optimal number of factors for each criterion.

With the above definition, we find that the optimal number of factors for single- and multi-region models is generally of the order of 10 and 15, respectively. As we have illustrated in Fig. 4, we have found that this number often closely matches the beginning of the model's precision plateau, thus preventing the model to overfit while maintaining a high precision.

Selection of assets for the calibration set and out-of-sample performance

Contrary to PCA-based models where all instruments described by the model must be part of the calibration set, the IPCA model is applicable to any instrument for which prices and valid characteristics exist. However, this raises an important issue: how do we ensure that the model will be good enough for instruments outside of the calibration set?

In the context of equity models, it is possible to include nearly all listed stocks in the calibration procedure. In this case, it is expected that instruments outside of the calibration set are less likely to exhibit features that are not already present in the calibration set. The training sample is fully representative of the whole opportunity set available to investors.

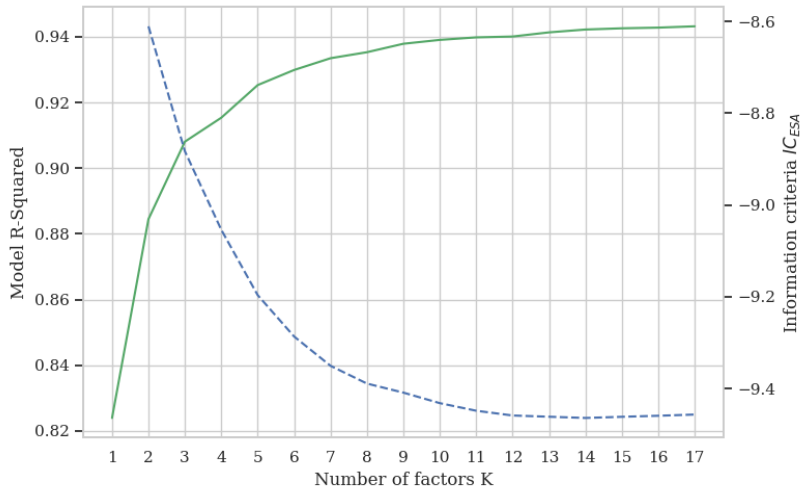

Figure 4: The optimal (lowest) point of the information criterion was found to match the starting point of the model's precision plateau well.

However, this strategy is not applicable to funds.There is in fact an infinity of possibilities to create a given portfolio from a large set of instruments, and only a fraction of all possible portfolios are actually invested in the form of publicly available instruments.

The fact that there exists an infinite number of possible portfolios does not mean that the calibration sample must also be infinite. In fact, in the context of calibrating models, it is necessary to separate the number of instruments and the number of return *patterns* (or *features*). A good calibration set must contain all of the features present in the investment universe, not necessarily all possible instruments. The addition of a further instruments exhibiting similar features does not enrich the calibration set. This is a particularly important point in the present context, as in practice funds actually exhibit fewer features than equities. Portfolios generally amplify commonly shared features among equities, while averaging out features found in only a couple of stocks. Also, most investable strategies tend to stay close to their benchmark, which further reduces the number of features found in funds strategies.

The problem of selecting instruments for the calibration procedure is two fold. First, the calibration set must be sufficiently large. Indeed, the precision of the model is linked directly to the size of the set; this is not specific to our model, but to statistical models in general. The convergence rate of many statistical models depends on the width and depth of the calibration set [Bai, 2003]. It was shown by Bai and Kelly that the convergence rate of the IPCA model is proportional to $\sqrt{N}$ for the statistical factors and $\sqrt{NT}$ for the gamma

matrix $\mathbf{\Gamma}$.

Second, the calibration set must contain as many features as possible, but with the fewest number of instruments. In order to illustrate to illustrate this claim, we estimate a large number of models using random calibration sets of different sizes. These sets include a fixed ensemble of one hundred funds, to which we added a random number of funds chosen randomly. We then measured the R-squared of each of these models on a test set containing only 100 random instruments outside of the calibration set.

The reason for choosing small test sets is to amplify the shortcomings of the calibration set: if a small test set contains even a few instruments with features not present in the calibration set, its R-squared will be strongly impacted. This is not the case in large test sets where the average R-squared of the model is unlikely to be affected by only a few instruments. The average of the model's out-of-sample precision for each sample size is shown in Fig. 5. The average of the in-sample precision is represented by the dashed line.

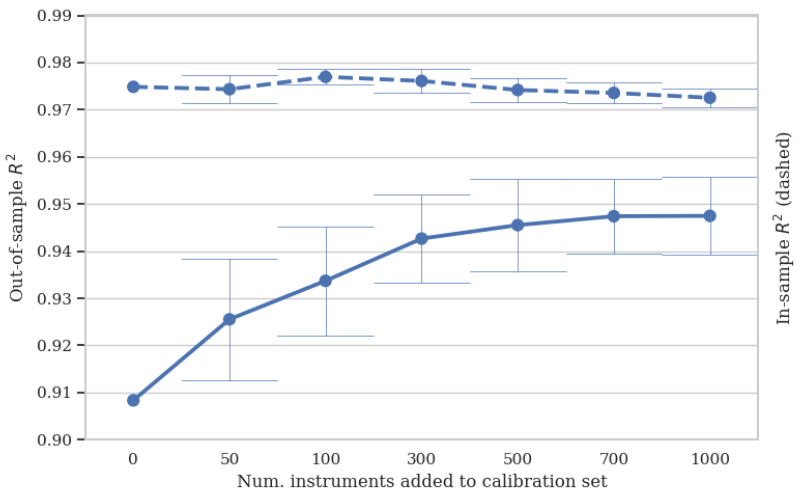

Figure 5: Out-of-sample precision of IPCA models calibrated on random calibration sets of different size. Each calibration set is formed by a fixed set of 100 publicly available funds to which a given number of funds chosen randomly is added. The precision of these models is tested on a large number of small random samples of instruments not present in the calibration sets. The average of the in-sample precision is represented by the dashed line.

What this test reveals is that past a certain number of calibration assets, the out-of-sample precision of the model reaches a plateau. It is worth noting that the number of assets for which this plateau is reached is not an intrinsic property of the model but depends on the investment universe itself. In our case, the out-of-sample precision of the model starts to plateau at around 300 instruments. This means that in this particular universe, any set of 300 or more instruments has a high likelihood of containing most of the features present in the investment universe.

The in-sample precision of the model (dashed line) illustrates the nature of the trade-off that has to be found: adding more instruments improves out-of-sample performance, but it also slightly reduces in-sample performance. As a consequence, the right number of calibration instruments should correspond to the number required to reach the out-of-sample precision plateau, but not much more.

Another issue illustrated in Fig. 5 concerns the actual choice of instruments for the calibration set. When the optimal number of calibration instruments is known, how should calibration instruments be selected so as to improve the model's out-of-sample performance? As it can be seen on the graph, calibration sets of equal size may leads to different out-of-sample performance. We know that the calibration set leading to the highest out-of-sample performance is also the set that contains the largest number of features. The creation of calibration sets with a high out-of-sample precision thus depends on the ability to evaluate the richness of a calibration set in terms of the number of features it contains.

The problem with quantifying the number of features present in a calibration set is difficult. However, we are helped by a strand of literature that addresses a problem that is conceptually similar to the selection of a good calibration sample, namely, the selection of test assets for the evaluation of linear factor models.

A common approach to comparing asset pricing models is to measure their relative efficiency at pricing the returns of a standard set of test instrument. The idea behind the utilization of test assets is that they should contain a large number of different return patterns so as to make their pricing as challenging as possible. The best model is thus the one that best reproduces the returns of the test assets (see e.g. [Cochrane, 2009] Chap. 12). The difference between the actual average return of an instrument and its estimate from the model is called "alpha". Accordingly, the lower the alpha among test assets, the better the model.

The construction of test assets has been the subject of an important research effort in the field of finance. Similarly to calibration assets, test assets need to possess the widest possible range of features and be representative of their investment universe. Their selection process and qualities are well-documented [Ahn et al., 2009].

The Gibbon, Ross and Shanken (GRS) test [Gibbons et al., 1989] is one of the most celebrated assessments of a model's goodness via the estimation of its alpha. Another important related quantity is the qualification of a model's inefficiency by [Hansen and Jagannathan,

1997] which is defined as

$$\text{Factor model inefficiency} \propto SR(f, A)^2 - SR(f)^2$$

Here the $SR(f, A)$ corresponds to the Sharpe ratio of the tangency portfolio associated with the model factors f and the test assets (A), whereas $SR(f)$ corresponds to Sharpe ratio the tangency portfolio associated with the models factors only. The tangency portfolio is the portfolio containing all instruments of a given investment universe whose weights maximize the Sharpe ratio.

What the above equation suggests is that if we form one tangency portfolio with the model's factors and another portfolio with the model and a number of test assets, then the difference in Sharpe ratios reflects the inefficiency of the model's factors to price the test assets. One way to reframe this result is to think in terms of diversification potential. All else being equal, the minimum variance of a portfolio is lowered by adding diversity to its investment universe. This is because a large diversity of returns provides many degrees of freedom for the optimization process to use in order to reduce variance while maintaining the same level of expected returns.

Because the number of features present in an investment universe and its diversification potential are conceptually very similar, the Sharpe ratio of its tendency portfolio thus provides an indirect estimation of the universe richness in terms of features.

Following the prescription of Bryzgalova [Bryzgalova et al., 2019] concerning the qualification of test assets, we formed 300 random calibration sets of different sizes and evaluated the Sharpe ratio of their *robust* tangency portfolios. Robust tangency portfolios are obtained by shrinking the expected returns of the test instruments towards their sample mean, which effectively makes them closer to a minimum variance portfolio[9]. Then we used these random calibration sets to estimate the parameters of risk models and measured the out-of-sample precision of these models on the remaining instruments.

[9] As it was shown by the authors in [Bryzgalova et al., 2019], the out-of-sample performance of sets formed using robust tangency portfolios with strong shrinkage was significantly higher than when using their non-robust counterparts.

We found that the value shrinking parameter did not strongly influence the results, and therefore set it to its maximum value. This means that the volatility σ_{CMV} of the minimum variance (MV) portfolio built using the calibration set should constitute a good proxy for the set's diversification potential. The calibration sets with the lowest variance are therefore expected to possess the largest number of features.

In order to examine the relationship between the volatility of the MV portfolio built using the instruments of the calibration set and the out-of-sample R-squared, we regressed the model's out-of-sample

R-squared R^2_{OOS} onto σ_{CMV}. We found the (negative) slope to be very significant, thus confirming the relationship between the richness of the calibration set and the diversification potential of the instruments it contains. The results are displayed in Table 4. We also estimated the relationship between the dispersion of the model's out-of-sample R-squared (R^2_{OOS}) and the minimum variance of the calibration set and found it equally significant. A low dispersion of R-squared means that the model's precision is not overly dependent on the exact composition of the calibration set. A scatter plot illustrating the relationship between σ_{CMV} and the out-of-sample precision of the model is shown in Fig. 6 for sets containing 700 random calibration assets.

	100	200	300	400	500	700
σ_{CMV} vs. R^2_{00S}	-0.01	-0.04	-0.17	-0.17	-0.22	-0.14
t-stat.	-0.62	-2.07	-4.05	-3.78	-4.88	-2.59
σ_{CMV} vs. $10^2 \times \mathrm{SD}(R^2_{00S})$	-7.55	-6.79	-6.76	-8.14	-8.74	-14.98
t-stat	-11.88	-11.10	-10.32	-10.18	-10.58	-15.70

Table 4: Relationship between the calibration sample minimum variance and the out-of-sample performance for different size of the calibration set. The values represent the slope of the line when regressing the σ_{CMV} against the model's R-squared out of sample R^2_{00S}.

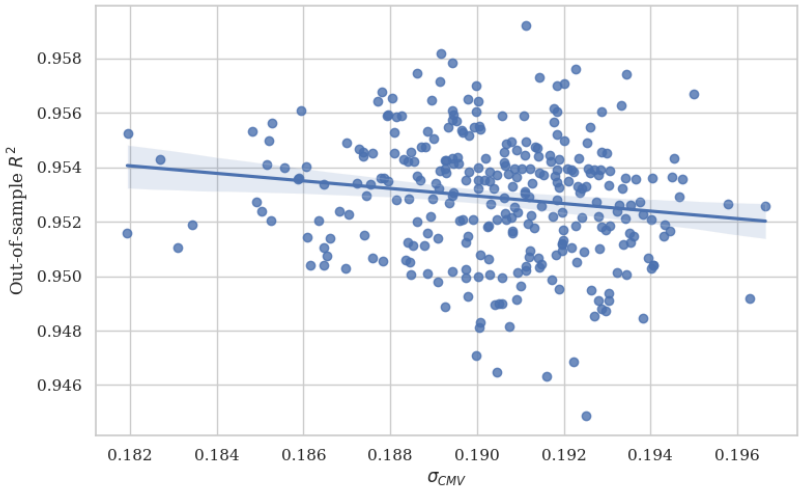

Figure 6: Out-of-sample precision of models calibrated on random calibration sets containing 700 instruments. The slope of the line is significant and confirms the relationship between the variance of the calibration set minimum variance portfolio and out-of-sample performance.

These results suggests that the optimal way of forming a calibration set is to select the instruments that most contribute to reducing the variance of a MV portfolio associated with a given investment universe. Such a calibration set increases the likelihood that models calibrated on this set will perform well out-of-sample. As we have seen before, number of instruments should also be sufficiently large so that the out-of-sample precision is already plateauing. This result is particularly important in the context of funds, as it is possible that

the set of publicly available funds may not be sufficiently large to ensure a good out-of-sample performance. In this case, it is advisable to complete the initial set of funds up to the optimal number of calibration instruments by adding e.g. a number of indices that are not necessarily invested but help increase the diversification potential of the calibration set.

10 Model precision and robustness

The precision of a risk model is of course an important criterion for its utilization in the context of risk analysis or an investment process. The measurement of precision is often performed via the model total or average R-squared over a given sample of instruments. However, aggregate measures of precision often do not provide a full picture with respect to a given model's performance. For this reason, we have performed a number of alternative measures in order to assess not only the model's average precision, but also how this precision was distributed within the sample, and whether it varied depending on market conditions.

Also, because the model is applicable to instruments that are not present in the calibration sample, we have tested whether the composition of the calibration sample impacted the model's precision on out-of-sample instruments.

Precision

A number of descriptive statistics regarding the precision of the IPCA model produced with the above characteristics and calibration procedures are shown in Table 5. These numbers were produced by calibrating and testing the model on the sample of US mutual funds described in Sec. 3. The number of factors was chosen to be optimal with respect to the information criterion developed in Sec. and is equal to $K = 11$. The R-squared of each instrument were subsequently measured and the universe average, alongside other statistics, are reported in Table 5. The 95% confidence interval is also reported. Here we report the results for the model where the characteristics are kept static. We found that overall the differences in precision between the static and dynamical version of the model were statistically insignificant.

For benchmarking purposes, we have performed the same precision measurements with concurrent models. We have included as benchmark both a PCA model with a similar number of factors,

and two multivariate time-series models. The first time-series model, labelled as "full", contains all the factors used to estimate the characteristics of the IPCA model. The second time-series model is the Fama-French five factors model (2x3) [Fama and French, 2015].

	$\bar{R}^2$	$\pm$ (95%)	Std(R^2)	Q1	Q4
PCA	0.97	0.00	0.05	0.96	0.99
IPCA-funds	0.95	0.01	0.08	0.95	0.99
Time-series (full)	0.88	0.01	0.10	0.82	0.96
Time-series (fama-5)	0.80	0.01	0.12	0.74	0.90

Table 5: Descriptive statistics of the R-squared distribution in the sample of US mutual funds for different risk models. The columns Q1 and Q4 correspond to the average of the values present in the first and last quartile of the distribution, respectively.

What is striking in Table 5 is the difference in precision between the IPCA model and the time-series model involving the same factors as those utilized to evaluate the model's characteristics. Given that these two models use the same risk-factors, the calibration of the cross-sectional model provides an advantage in terms of precision. The systematic superiority of cross-sectional over time-series models, in spite of the qualitative similarity of their inputs, was already noticed and documented in [Fama and French, 2015].

Although the IPCA has a significantly larger precision on average than the full time-series model, the cumulative distribution of the R-squared within the sample reveals that time-series models tend to have much larger tails. This is illustrated in Fig. 7. By larger tails we means that for a small fraction of funds, time-series model tend to perform very poorly. In this respect, the consistency of the IPCA is much closer to that of PCA models, in spite of its fundamental nature.

Another form of consistency is concerned with market conditions. In Fig. 8 we have selected at random samples of 52 weeks and reported the different models precision and the sample volatility. The slope of the lines represent the linear relationship between the models precision and the sample's volatility. The positive slope is expected as all models calibration procedures aim at reducing the differences between the observed and the model returns. Because these differences become larger when volatility is high, all calibration procedures will prioritize high-volatility phases. What we observe again is that the PCA and IPCA model exhibit less sensitivity to the volatility of the sample and hence deliver more consistent performance.

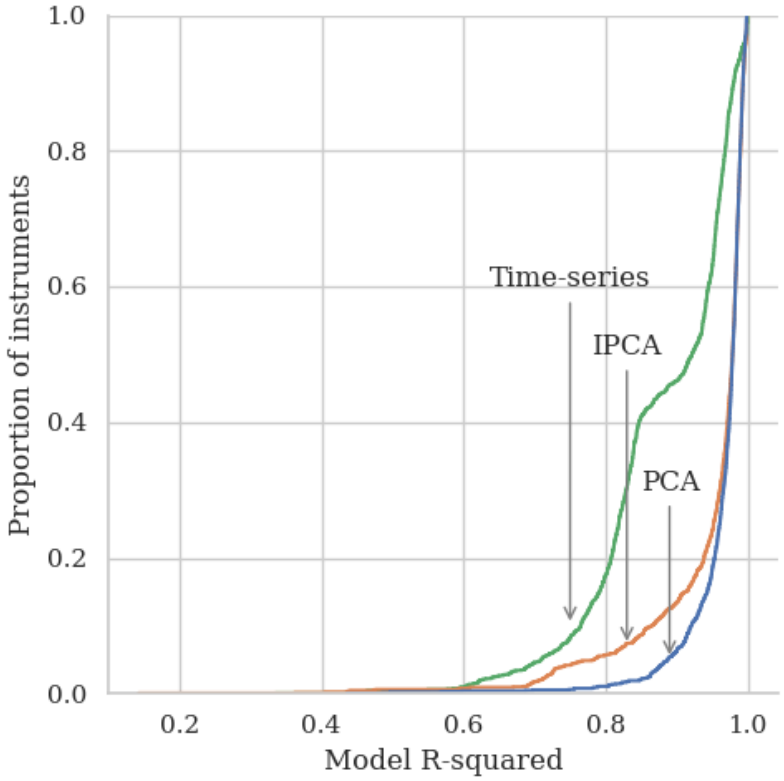

Figure 7: Proportion of funds with precision below the number shown on the x-axis. For the same instrument, the precision of the time-series model is generally lower than its counterparts.

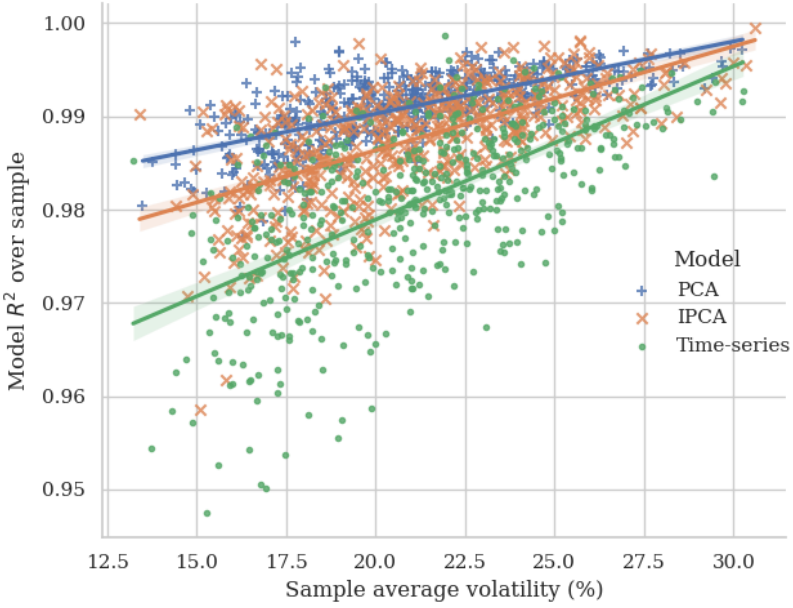

Figure 8: The precision of various models depending on the sample volatility.

Use of the model for stocks

So far we have focused on the performance of the model for describing the returns of funds. Logically, the calibration set should contain most of the returns patterns that exist in the wider population, and therefore should consist of a large number of funds.

As we have already explained, the IPCA model allows instruments that are not present in the calibration set to be explained by the model. Because funds are themselves made out of equities, it interesting to ask whether a model calibrated on funds is able to describe the returns of equities with a precision on par with models that would have been calibrated on equities. This amounts to asking whether stocks exhibit returns patterns that are in fact not very different from those of funds.

The applicability of a funds model to equities would have several advantages, especially in the context of risk and performance decomposition. With only one model, it would be possible to coherently analyse the performance of a portfolio starting from the high level of individual funds, down to the contribution of each equity when this information is available. By coherently, we mean here that the contributions of both stocks and funds will sum to the total risk of the portfolio as measured by the model.

In order to evaluate the applicability of a model calibrated on funds to stocks, we have compared the performance of three models. The first is an IPCA model calibrated on a sample of US funds. It is the same model whose performance is analyzed in Sec. 10. The second model is an IPCA model that has been calibrated on stocks. To do so we have used a sample of 2000 US stocks and evaluated their characteristics using the same risk factors, and methodology, as those we used for funds. Lastly, we have evaluated a statistical model using the principal components decomposition of the stocks returns over a trailing period of five years. All models have the same number of systematic factors, and have been evaluated each month from the first of January 2015 until December 2022, or a total of 96 months.

In Figure 9 we report the precision of the different models. The R-squared of each stock was evaluated each month using the three models, which represents a total of 162,940 values of R-squared for each model. Two observations are in order here. First, the hierarchy of performance between the models is quite expected, as the model calibrated on funds is less precise than the two other models calibrated on stocks. However, what is less expected in these results is actual size of these differences in precision: it is indeed very small.

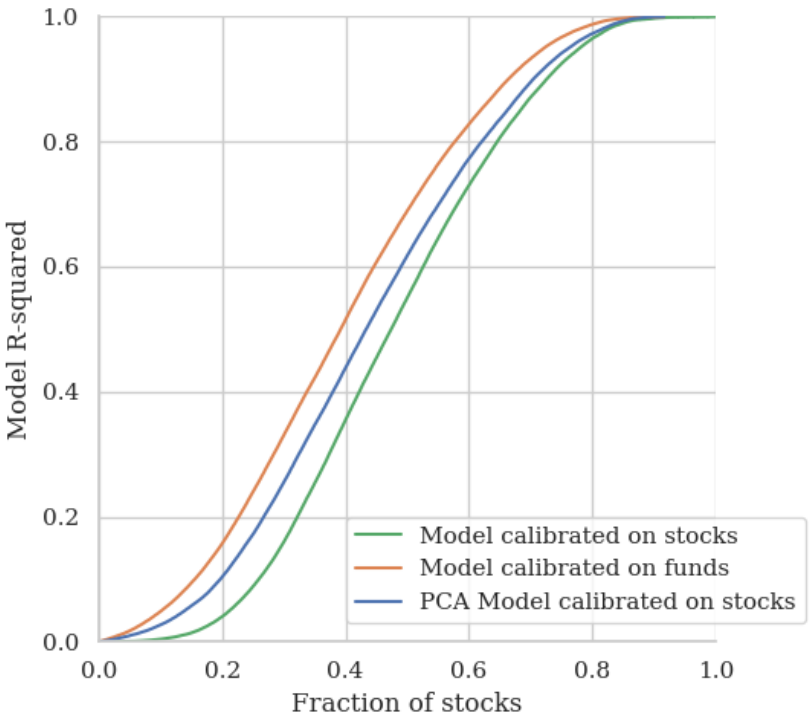

Figure 9: The precision of various models applied to a sample of 2000 US stocks between January 2015 and December 2022.

The difference of the average R-squared between the IPCA model calibrated on funds (40%) and the model calibrated on stocks (44%) is a little bit less than 4%. For comparison, the R-squared of the IPCA model calibrated on stocks is also 3-4% lower than the PCA model, although the later cannot be used to decompose risk into fundamental drivers. Note that for larger stocks, such as those present in the S&P 500, the difference in R-squared between models is a lot smaller, of the order of 2%, that is 47%, 49% and 51% for the IPCA models calibrated on funds, stocks, and the PCA model, respectively.

In order to illustrate what a difference of 4% in R-squared represents in terms of risk, let us consider stocks with a volatility of about 20%. A model with a R-squared of 45% means that about 13.5% of these stocks total volatility is explained by the model's factors. A R-squared of 40% means that the explained volatility for these same stocks is about 13%. So the difference in R-squared between the two models is actually small, even unimportant from a practical point of view. So although the model calibrated on funds is slightly less precise, it does not differ much, from a practical point of view, from an IPCA model calibrated on stocks. However, its high precision for both funds and equities makes it possible to coherently analyze portfolios of both funds and equities.

In Sec. 11 we will examine how these differences in R-squared at the individual stock level affects the models biases and their ability to forecast risk over a given time horizon.

11 *Model applications*

The main application of riskmodels that we are going to cover in the next paragraphs is the decomposition of the portfolio's volatility and returns. Lastly, we discuss the application of the model to portfolio optimisation from the angle of its relative stability.

Using the definition of a portfolio's returns $r_{w,t}$ from time $t-1$ to t from Eq. 2, we obtain the usual approximation of the portfolio's volatility as:

$$\sigma_w = \sqrt{\boldsymbol{w}^T(\boldsymbol{\Omega}_{sys} + \boldsymbol{\Omega}_\varepsilon)\boldsymbol{w}} \tag{9}$$

where:

$$\boldsymbol{\Omega}_{sys} = \frac{1}{T}\sum_t \boldsymbol{c}_t \boldsymbol{\Omega}_{f,t} \boldsymbol{c}_t^T = \sum_t \boldsymbol{\Omega}_{sys,t},$$
$$\boldsymbol{\Omega}_\varepsilon = \frac{1}{T}\sum_t \mathrm{diag}(\varepsilon_t^{*2})$$

Here the matrix $\boldsymbol{\Omega}_{f,t} = \boldsymbol{\Gamma}\boldsymbol{f}_t\boldsymbol{f}_t^T\boldsymbol{\Gamma}^T$ is $(L \times L)$ and $\boldsymbol{c}_t$ is the $(N \times L)$ matrix containing the characteristics of the N instruments.

The dynamical nature of the exposures $\boldsymbol{c}_{i,t}$ adds an extra dimension to the definition of the covariance matrix compared to static models. The above formula is thus similar to a sum of *point-in-time* covariance matrices.

Decompositions of risk and performance into contributions associated with the instrument characteristics are obtained via the application of the *Euler decomposition*. In Appendix C, we provide a derivation of this important formula, and explain why it is applicable to the case of risk and return analysis.

By writing both risk and rewards as functions $f(\boldsymbol{c}_t)$ of the characteristics over a given period of time, the Euler decomposition gives:

$$f(\boldsymbol{c}_t) = \sum_t \sum_i \sum_\ell \underbrace{\left(\frac{\partial f(\boldsymbol{c}_t)}{\partial \boldsymbol{c}_t} \circ \boldsymbol{c}_t\right)_{i\ell}}_{\substack{\text{contribution of the} \\ \ell\text{th characteristic of the} \\ i\text{th instrument at time } t}} \tag{10}$$

where the symbol $\circ$ represents here the Hadamard product between two matrices.

As we are going to illustrate in the next sections, the IPCA model lends itself well to such decomposition and hence provides a high level of granularity for risk and performance analysis.

Risk decomposition

For decomposing risk at each time t, the function $f(\boldsymbol{c}_t)$ becomes the portfolio volatility σ_w. The decomposition of total risk using Eq. 10 gives:

$$\sigma_w = \sum_t \sum_i^N \sum_\ell^L (RC_t)_{\ell,i} + \sum_t \sum_i^N (RC_t)_{\varepsilon,i}$$

where the risk contribution RC_t is an $L \times N$ matrix whose elements $(RC_t)_{\ell,i}$ correspond to the *point in time* contribution of the ℓ^{th} characteristic of the i^{th} instrument to the total risk and $\boldsymbol{w}_t$ is the $(N \times 1)$ vector of portfolio weights at every point in time.

As we have seen in the previous section, the contributions to risk are defined as:

$$(RC_t)_{\ell,i} = (\boldsymbol{c}_t \circ \frac{\partial \sigma_w}{\partial \boldsymbol{c}_t})_{\ell,i}$$
$$(RC_t)_{\varepsilon,i} = (\boldsymbol{\varepsilon}_t \circ \frac{\partial \sigma_w}{\partial \boldsymbol{\varepsilon}_t})_i$$

The derivative $\frac{\partial \sigma_w}{\partial \boldsymbol{c}_t}$ and $\frac{\partial \sigma_w}{\partial \boldsymbol{\varepsilon}_t}$ are obtained using a number of identities common in matrix derivatives calculus [Petersen and Pedersen, 2012]:

$$\begin{aligned}\frac{\partial \sigma_w}{\partial \boldsymbol{c}_t} &= \frac{1}{2}\frac{1}{\sigma_w}\frac{1}{T}\frac{\partial}{\partial \boldsymbol{c}_t}\left[\boldsymbol{w}_t^T \boldsymbol{c}_t \boldsymbol{\Omega}_{f,t} \boldsymbol{c}_t^T \boldsymbol{w}_t\right] \\ &= \frac{1}{\sigma_w}\frac{1}{T}\boldsymbol{W}_t \boldsymbol{c}_t \boldsymbol{\Omega}_{f,t}\end{aligned}$$

where $\boldsymbol{W}_t = \boldsymbol{w}_t \boldsymbol{w}_t^T$. The above formula is also useful when evaluating the confidence interval around the estimate of the volatility using the delta method. In this case, the confidence intervals stems from the uncertainty associated with the value of the characteristics. As we have mentioned before, the estimation of the characteristics is based on regressions, and therefore it has a standard error associated with it. Even when characteristics are not based on regressions, their value is always subject to uncertainty, and so the estimation of a confidence interval around the estimate of the volatility is in principle always possible. This interval is particularly relevant when comparing risk between strategies.

The derivative of the volatility with respect to the specific risk is simpler:

$$\begin{aligned}\frac{\partial \sigma_w}{\partial \boldsymbol{\varepsilon}_t} &= \frac{1}{2}\frac{1}{\sigma_w}\frac{1}{T}\frac{\partial}{\partial \boldsymbol{\varepsilon}_t}\boldsymbol{w}_t^T \mathrm{diag}(\varepsilon_t^{*2})\boldsymbol{w}_t \\ &= \frac{1}{\sigma_w}\frac{1}{T}\mathrm{diag}(\boldsymbol{W}_t)\mathrm{diag}(\varepsilon_t)\end{aligned}$$

Note that summing all risk contributions over every dimension (times and instruments) returns the total volatility of the portfolio.

For single funds, the risk decomposition simplifies considerably and is obtained by applying the following modifications to the above formulas:

$$\begin{aligned}\boldsymbol{w}\boldsymbol{w}^T &\to 1 \\ \boldsymbol{c}_{i,t} &\to \boldsymbol{c}_t\end{aligned}$$

The ability to map the risk contributions back to the characteristics allows us to separate the exact risk contributions coming from rewarded and unrewarded factors. For instance, the contribution of the risk associated with exposures to the "Value" risk factor is simply given by:

$$\sum_{i,t}(RC_t)_{i,\ell=\text{Value}}$$

As an illustration, the average contributions to total risk associated with each risk factor (exempt market risk) across our sample of US mutual funds are shown in Fig. 10. Note that the average contribution of market risk is around 18%. The specific risk is denoted as "other" on the graph. Factor exhibiting a negative average contribution to risk more often diversify risk than contribute to it.

In Fig. 11, the average factor contribution to total risk is shown for a sub-sample of funds that contain the word "Value" in their name. The contribution to risk from the different factors for this sub-sample is consistent with expectations. We have observed that the variance of risk exposures among funds with clearly stated objectives is often surprisingly low. This suggests that funds with similar objectives, even if they are not factor-based such as e.g. "High-dividend" funds, have very similar exposures not only to one or two factors, but to almost all factors.

Another dimension that is often overlooked is the historical one. Grouping risk contributions not only by factor types (rewarded or

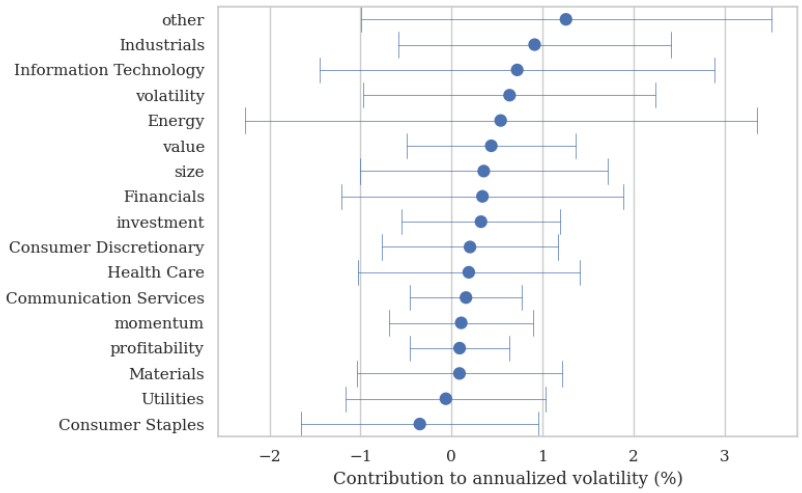

Figure 10: Average (point) and standard deviation (bar) of risk contributions per risk factor across our sample of US mutual funds.

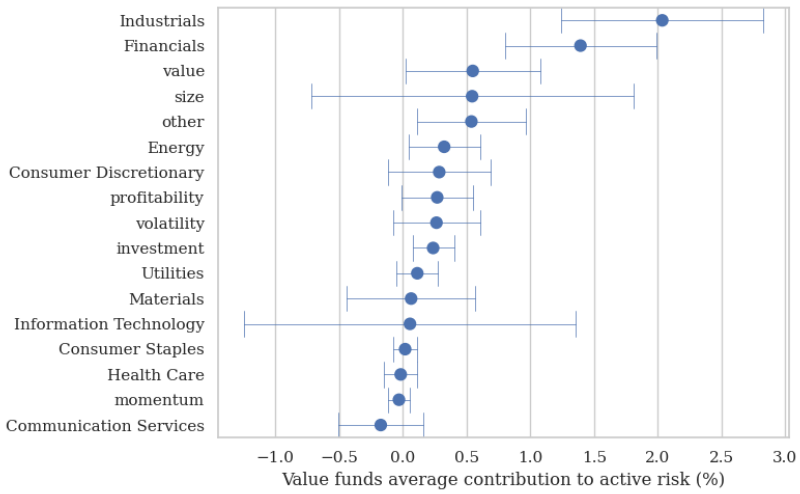

Figure 11: Average (point) and standard deviation (bar) of the factors contributions to total risk associated for a sub-sample of funds containing the word "Value" in their name.

unrewarded) but by time periods provides an insight into the realization of risks. Oftentimes, the largest part of the total risk comes from the contributions of a few short periods of time. Risk tends to appear in bursts, which the above representation is capable of identifying. This can be observed, for instance, in Fig. 12 where the sample average quarterly contributions to total risk associated with the "Value" characteristics are shown .

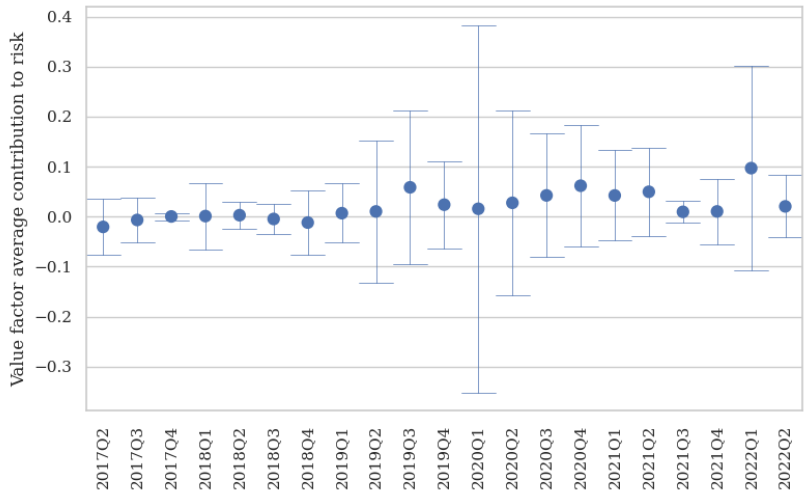

Figure 12: Average (point) and standard deviation (bar) of the quarterly contributions to total risk associated with the "Value" characteristics across our sample of US mutual funds.

Returns decomposition

In addition to decomposing risks, the risk model also allows an exact decomposition of systematic returns via the Euler formula. For a portfolio $\boldsymbol{w}$ the decomposition of the cross-section of returns is particularly straightforward, as:

$$\begin{aligned} r_{w,t} &= \boldsymbol{w}^T \boldsymbol{c}_t \boldsymbol{\Gamma} \boldsymbol{f}_t + \varepsilon_t \\ &= \sum_{\ell} \sum_{i}^{N} (\boldsymbol{r}_{w,t})_{\ell,i} + \varepsilon_t \end{aligned}$$

where the term $(\boldsymbol{r}_{w,t})_{\ell,i}$ is the contribution of the i^{th} instrument's ℓ^{th} characteristic to the weekly return obtained using Eq. 10 where:

$$(\boldsymbol{r}_{w,t})_{\ell,i} = (\boldsymbol{c}_t \circ \frac{\partial \boldsymbol{r}_t}{\partial \boldsymbol{c}_t})_{\ell,i} = (\boldsymbol{c}_t \circ \boldsymbol{w} \boldsymbol{f}_t^T \boldsymbol{\Gamma}^T)_{\ell,i}$$

Granular performance attributions are obtained by aggregating the individual contributions along the instrument and characteristic axes.

For instance, the performance *at time t* attributable to the portfolio exposure to the quality factor is given by:

$$\sum_i (\boldsymbol{r}_{w,t})_{\text{quality},i}$$

At each point in time, it is thus possible to determine the contribution to performance coming from each risk factor. However, contrary to risk decomposition, the contributions of each risk factor to cross-sectional returns cannot be summed to evaluate the contributions *over time*. This is a consequence of returns being compounded rather than summed to measure performance over time:

$$r_w = \prod_t (1 + r_{w,t}) - 1 \neq \sum_t r_{w,t}$$

When point-in-time returns are very small, the sum over time might provide an approximation of the period performance, but as the number of time periods grows, this approximation becomes poor.

Because the ability to sum contributions both historically and cross-sectionally is very desirable to provide insights into the drivers of performance, several methods have been developed in order to "link" the cross-sectional returns. By linking, we mean that the values of cross-sectional returns are adjusted so that the sum corresponds to the exact performance over a given period of time. Linking techniques amount to finding an adjustment factor γ_t so that:

$$r_w = \sum_t \tilde{r}_{w,t} \quad \text{with} \quad \tilde{r}_{w,t} = \gamma_t r_{w,t}.$$

As it can be seen from the above formula, once adjusted, the daily contributions can be summed over time which allows the decomposition of performance over time, characteristics and instruments.

The downside of linking methodologies is that they distort cross-sectional returns for the sake of additivity. We use a linking methodology known as the Frongello method [Frongello, 2002] that was recently proven to mitigate the problem of amplifying small returns that occurs with the well-known Cariño method (see e.g. [Jiang and Saenz, 2014]). Frongello adjustments are defined as:

$$\gamma_t^F = \left[\prod_{i=1}^{t-1} (1 + r_{w,t}) \right]$$

Once returns are properly adjusted, a granular, exact performance decomposition over time and across instruments is again possible.

Note that the annualization of returns is obtained in a similar way by multiplying the return contributions by a coefficient (see Appendix D).

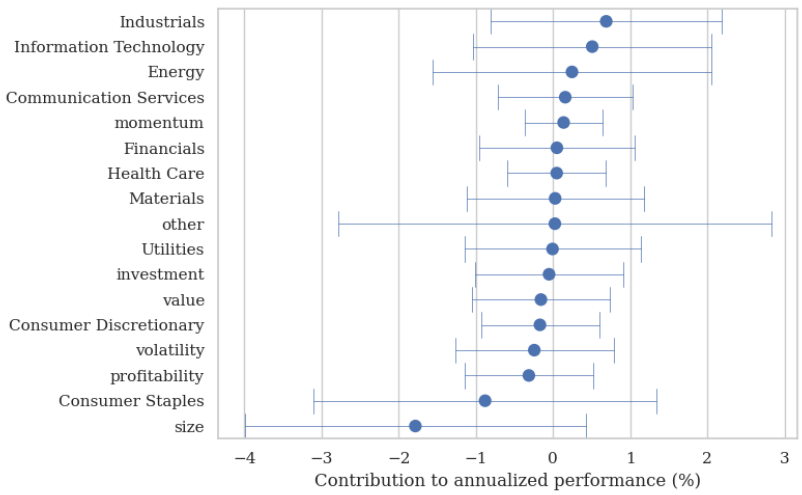

Figure 13: Average (point) and standard deviation (bar) of performance contributions per risk factor across our sample of US mutual funds.

For illustration, the average contribution to performance over the period considered is shown in Fig. 13. Also, in order to display the ability to visualize the variations of risk contributions over time, the average quarterly contribution to performance associated with the "Value" characteristics is shown in Fig. 14. Over most of the period, the contributions to performance of the "Value" factor is almost zero for most funds, whereas its most significant contributions come in burst over a couple quarters. The historical dimension of the contributions to risk thus provide context to the aggregate figure and often exemplifies the difficulty of timing factor performance. Finally, note that for this sample of large US mutual funds, the performance that is not explained by factors is statistically indistinguishable from zero, a feature that was observed and documented in many other places.

Perspective on the IPCA model for portfolio optimisation

Beyond the ability of the model to provide insights into a portfolio's drivers of risk and performance, what we are going to show now is that the IPCA model is also an objectively excellent candidate for portfolio optimisation.

In this context, we are going to examine here a number of key properties for assessing the adequacy of the model for portfolio optimisation. The estimation of a portfolio's weights using a risk model

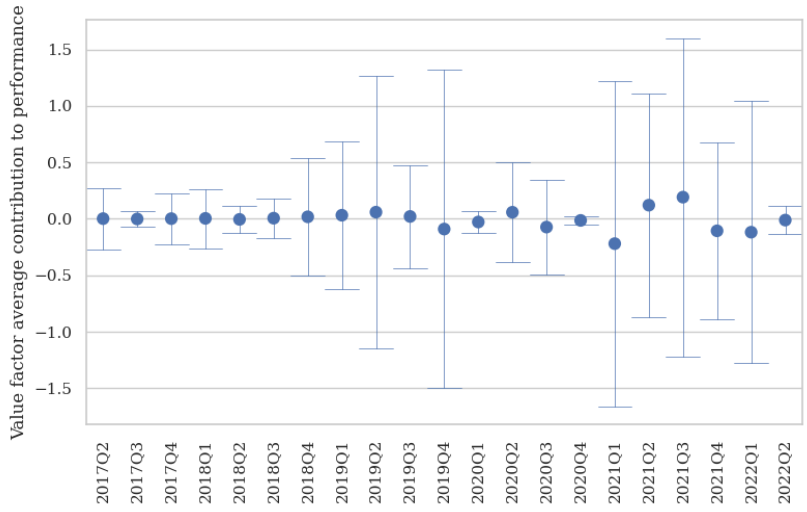

Figure 14: Average (point) and standard deviation (bar) of the quarterly contributions to performance associated with the "Value" characteristics across our sample of US mutual funds.

requires essentially two types of qualities. First the portfolio composition should not be overly affected by small changes in the model's covariance matrix. That is, the portfolio should be stable to small changes caused by, e.g. a few missing prices. This property is generally refered to as the stability of a model. It is practically important in the context of portfolio optimisation, since unstable matrices have the potential to unnecessarily increase the portfolio's turnover. Therefore the stability of a covariance matrix may have a financial impact.

Second, a portfolio estimated today should be based on information that will be true not only today, but possibly until the next rebalancing date. In other terms, the model should not have strong biases, and provide relatively good forecasts. When using the matrix to reduce tracking error for instance, the targeted tracking error measured with today's covariance matrix should possibly persist until the next rebalancing date.

The estimation and comparison of the IPCA model stability, biases and forecastablility are the topic of this section.

Although we will develop this topic elsewhere, we have also found that factors orthogonality combined with the knowledge of the characteristics confidence interval makes the IPCA model an ideal candidate for the implementation of robust portfolio optimisation, as defined e.g. [Goldfarb and Iyengar, 2003]. Indeed, the factors orthogonality greatly simplifies the implementation of these techniques, while the estimation of error ranges is straightforward. The treatment of this topic goes beyond the intent of this paper, but this property is also worth mentioning since robust optimisation is often overlooked

for its difficulty of implementation.

Stability

The stability of a model in the context of portfolio optimisation quantifies how a small change in the covariance matrix translates into changes in the portfolio weights. Small changes in the covariance matrix of a stable model should not really change the weights of a portfolio that has been evaluated using it.

It is useful to start by defining what do we mean by a portfolio that has been optimised using a covariance matrix. In fact, most optimisation programs with some form of risk reduction in their objectives involve solutions that depend on the covariance matrix to be inverted[10]. So the possibility to invert the model's covariance matrix is in fact a prerequisite for using it in a portfolio optimisation program. A simple illustration of this relationship is the unconstrained minimum variance portfolio,

[10] Examples of objectives involving risk are e.g. mean-variance minimisation, tracking error reduction or diversification maximisation.

$$\boldsymbol{w}^{\star}_{MV} = \boldsymbol{\Omega}^{-1}\mathbb{1}.$$

which is a solution to the simple risk minimisation program:

$$\boldsymbol{w}^{\star}_{MV} = \arg\min \boldsymbol{w}^{T}\boldsymbol{\Omega}\boldsymbol{w}.$$

In this simple case, the minimum variance portfolio is similar to the solution of a linear system of equations defined by $\boldsymbol{\Omega}$. This simple examples illustrates the similarity between the stability of a system of equations and that of a covariance matrix in the context of portfolio optimisation. The conditions for a system of equations to be stable have been studied extensively, and they directly apply to the problem of measuring the stability of a risk model. When the covariance matrix cannot be inverted, the minimum variance portfolio shares the same fate as the solutions to any ill-conditioned linear systems[11], it remains undefined. Another problem affecting minimum variance portfolios that share similarities with ill-defined linear systems is the sensitivity of the portfolio to small changes in the covariance matrix. When small changes in the elements of the covariance matrix cause large changes in the optimal portfolio, the problem is said to be unstable.

[11] Linear systems of equations defined as $Ax = y$ are solved by inverting the matrix A, that is $x^{\star} = A^{-1}y$.

A useful metric that was developed for the study of linear systems that captures the sensitivity of the optimal portfolio to small changes in the covariance matrix is called the *condition number*, which is often denoted by κ. For a good introduction to the condition number, see e.g. [Belsley, 1991]. The condition number is proportional to the impact on the portfolio's weight when the elements of the covariance

matrix are changed by a small amount. So when the condition number is high, the stability of the problem is low. When the covariance matrix is singular or ill-defined, then a small change in the covariance matrix causes the solution to explose. This is why the *distance to singularity*, which simply corresponds to the inverse of the condition number, is often used instead. It is also an easier metric to deal with: when the distance to singularity is low, the stability of the solution is also low. A distance of zero indicates that the problem is ill-conditioned.

The condition number of a matrix is proportional to the ratio between its highest and the lowest non-zero singular values, or eigenvalues. The evaluation of a covariance matrix eigenvalues, which is easily obtained, thus provides an immediate insight into its stability. This definition of stability also means that the value of the smallest eigenvalue has an important impact on its stability. An eigenvalue close to zero means total instability. Therefore one way of increasing the stability of a covariance matrix is to find a close equivalent whose smallest eigenvalue is as large as possible. This is done usually by evaluating the covariance matrix singular-value decomposition

$$\mathbf{\Omega} = \boldsymbol{U}\mathbf{\Sigma}\boldsymbol{U}^{\boldsymbol{T}} \text{ with } \mathbf{\Sigma} = \begin{pmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_N \end{pmatrix}$$

and set its smallest singular values, starting from σ_N, to zero. Doing so indeed creates an approximation of the covariance matrix. The difference between the covariance matrix and its approximation will remain small if the singular values that are set to zero are small. This is the idea behind many PCA-based models risk models. The effective number of factors in such models is said to correspond to the number of non-zero singular values. By only keeping factors associated with large eigenvalues, the stability of the model increases dramatically. This is why PCA-based models are often chosen for portfolio optimisation tasks.

For illustration purposes, we have estimated the SVD decomposition of our funds sample covariance matrix for different number of factors. Their associated distance to singularity is shown In Fig. 15. It is well known that covariance matrices become singular when $N > T$ (see e.g. [**?**] and references therein) and accordingly we find that the sample covariance matrix becomes singular for $K = T$. Note that for numerical application, the zero is located around 10^{-14}. This means that for most numerical application, solutions will not be available for a number of factors $K < T$. As it is generally the case, the distance to singularity decrease exponentially with the number of factors.

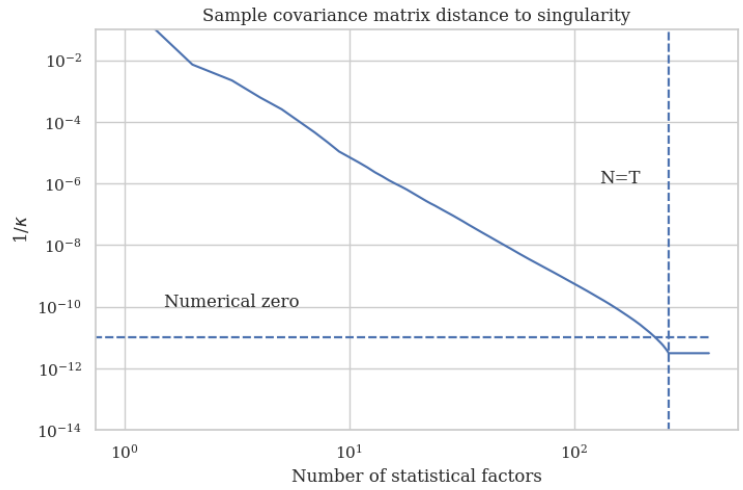

Figure 15: Distance of the sample covariance matrix to singularity, as measured by its condition number.

In the case of the IPCA model, the number of non-zero singular values corresponds to the number K of statistical factors. For illustration purposes, in Fig. 16 we show the distance to singularity of a model calibrated on our sample of US mutual funds for different values of K. What is interesting to note here is that between a number of e.g. 10 factors to about 20-30 factors, the distance to singularity decreases by two orders of magnitude.

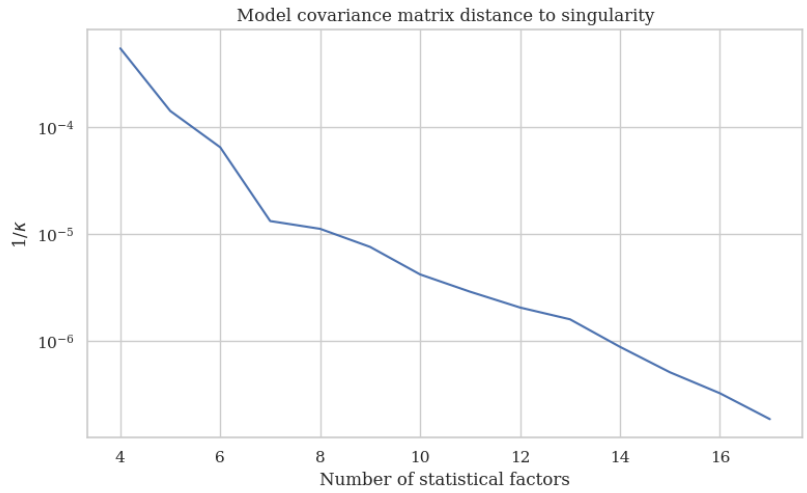

Figure 16: Distance of the IPCA model's covariance matrix to singularity. The model is calibrated on a sample of 400 US mutual funds.

As we have shown before, the number of statistical factors is selected so that the model is able to maintain a level of precision comparable to what it would have if the number of factors was set to its

maximum value.

When the IPCA is calibrated with its maximum number of factors, it in fact corresponds to the standard cross-sectional model. It is typical for single- or multi-regional cross-sectional models to involves a few tens of variables. What is shown in Fig. 16 is that the dimensionality reduction mechanism of the IPCA model, makes it more suitable for risk-based portfolio optimisation than standard cross-sectional models. As the number of factors passes from a few tens to about 10 to 15, the stability of the covariance matrix, as measured by its distance to singularity, improves by almost two orders of magnitude.

Forecastability and biases

Beyond the stability of the covariance matrix, another key concern for portfolio optimisation is the so called bias of the matrix. The bias relates to the property of the risk model not to systematically over- or underestimate the risks over the next period. Also, a model that would switch between over- and underestimating risk could not be considered as very reliable for portfolio optimisation. Therefore the volatility of the bias provides information about the stability of the model's forecasting quality.

Both the bias and its volatility quantify the ability of the model, and consequently the portfolios optimised with it, to maintain their properties until the next rebalancing date. The bias is thus a measure that takes the model prediction and compares it with the realized risk. Here we will use the work of Briner *et al.* in [**?**] which provides the following definition for the bias:

$$\begin{aligned} z_w(t) &= \frac{r_w(t)}{\sigma_w(t-1)} \text{ with } \sigma_w(t) = w^T \Omega_t w \\ b_w(T) &= \sqrt{(1/T)\sum_{t=1}^{T}(z_w(t) - \bar{z}_w)^2} \end{aligned} \tag{11}$$

where the covariance matrix evaluated at time t is denoted by Ω_t and the portfolio's return over the period ending at time t by $r_w(t)$. A common issue when evaluating realized risk is that the time horizon between rebalancing dates is often short. This only allows a noisy measurement of the realised risk, which in turns affects the precision of biases estimated with it. This problem is exacerbated when the model and realized risk are measured with returns at different frequencies. The above definition is in fact very well-thought as it solves this important problem. Instead of estimating the realized risk over the subsequent period, it assumes that the ratio of the volatility at time t and subsequent return should converge to one if the model

does not have a bias. Note that here we have considered rebalancing periods of one month, which is typical when testing risk models.

We have evaluated three different covariance matrices each month between January 2015 and December 2022. The first is simply a sample covariance matrix estimated over the past 5 years starting from the rebalancing date. This sample covariance matrix was normalized using the method of [**?**] to avoid singularities. The two other covariance matrices are obtained from an IPCA model calibrated using the characteristics of US funds for the first model, and stocks for the other. The bias statistic is then measured on random portfolios of stocks with different sizes. We have considered portfolio sizes of 20,50, 100 and 200 stocks and generated 100 equally weighted (EW) and capitalisation weighted (CW) portfolios for each size, or 800 portfolios in total. Over the 96 months between January 2015 and December 2022, we have therefore estimated 76,800 different values of $z_w(t)$ to build the statistics shown below. Also, for each portfolio, we have estimated its tracking error against a capitalisation weighted portfolio containing all the stocks present at each time. This statistics provides information about the model's forecastability for tracking error.

Portfolio Type	Size	Ledoit-Wolf	IPCA funds	IPCA stocks
EW	20.00	1.23	1.13	1.14
	50.00	1.34	1.22	1.22
	100.00	1.36	1.24	1.24
	200.00	1.40	1.27	1.27
CW	20.00	1.14	1.08	1.08
	50.00	1.19	1.12	1.12
	100.00	1.23	1.16	1.14
	200.00	1.27	1.17	1.16

Table 6: Biases of volatility forecasts estimated between January 2015 and December 2022.

It is important to note that the portfolios are generated randomly. In some cases, biases statistics are calculated on e.g. sector or factor portfolios, which normally favours models built around similar risk factors. The biases estimated using random portfolio are thus expected to be more conservative. The average of volatility biases per portfolio size and type are shown in Table 6, while average biases for tracking error measurements are shown in Table 7.

What this measurements show is in fact quite surprising. In fact, the calibration universe of the model, be it on stocks or funds, does not greatly affects neither biases nor their volatility. We have seen in Sec. 10, the precision of the model calibrated on stocks to describe

Type	Size	Ledoit-Wolf	IPCA funds	IPCA stocks
EW	20.00	1.17	1.07	1.06
	50.00	1.27	1.15	1.11
	100.00	1.30	1.18	1.12
	200.00	1.32	1.18	1.12
CW	20.00	1.03	1.01	1.01
	50.00	1.03	1.03	1.00
	100.00	1.04	1.06	1.01
	200.00	1.04	1.05	0.99

Table 7: Biases of tracking error forecasts estimated between January 2015 and December 2022.

single stocks is higher than that of the model calibrated on funds. However, when stocks are aggregated into a portfolios, the two models perform equally well to describe the risks of these portfolios.

It is commonly known that the bias of a portfolio is reduced by a more reactive measurement of volatility. This is why a number of risk models destined to be used for portfolio optimisation actually replace the diagonal of the covariance matrix by a more contemporaneous measure of volatility. We have implemented a variation of the above models that involves more reactive volatility estimates by proceeding as follows. For the Ledoit-Wolf (LW) sample covarance matrix, we have replaced the diagonal elements by the corresponding EWMA estimate of the instruments volatility, that is:

$$(\Omega_t^{LW})_{ii} \overset{\text{reactive}}{\longrightarrow} (1-\lambda)\sum_{n=0}^{\infty}\lambda^n r_{t-n}^2$$

For IPCA models, the covariance matrix depends on the volatility of the factors (see Eq. (9))

$$\mathbf{\Omega}_{f,t} = \mathbf{\Gamma}\boldsymbol{\sigma}_{f,t}\mathbf{\Gamma}^T \tag{12}$$
$$\boldsymbol{\sigma}_{f,t} = \boldsymbol{f}_t\boldsymbol{f}_t^T \tag{13}$$

and so in order to make the covariance matrix of this model more reactive we have replaced the diagonal of the the matrix $\boldsymbol{\sigma}_{f,t}$ by the EWMA estimates of the factors $\boldsymbol{f}_t$ variances

$$(\boldsymbol{\sigma}_{f,t})_{ii} \overset{\text{reactive}}{\longrightarrow} (1-\lambda)\sum_{n=0}^{\infty}\lambda^n f_{i,t-n}^2$$

The reactivity parameter λ for both the sample and IPCA models we set with a half-life of 6 months. The results shown in Tables 8 and

9 were obtained using these modified models. The same portfolios were used to estimate the $z_w(t)$ using both the previous risk models and their more reactive adjustments.

The volatililty adjustments did actually drastically improved the model biases. We also again observe very little differences between the IPCA models calibrated on funds or stocks. The IPCA models tend to have lower biases than the sample covariance matrix, probably as a result of their better stability.

Portfolio Type	Size	Ledoit-Wolf*	IPCA* funds	IPCA* stocks
EW	20.00	1.21	1.03	1.04
	50.00	1.32	1.08	1.09
	100.00	1.35	1.08	1.09
	200.00	1.39	1.10	1.11
CW	20.00	1.12	0.98	0.99
	50.00	1.18	1.01	1.00
	100.00	1.22	1.03	1.01
	200.00	1.25	1.04	1.02

Table 8: Biases of volatility forecasts estimated between January 2015 and December 2022 with reactive covariance matrices.

Portfolio Type	Size	Ledoit-Wolf*	IPCA* funds	IPCA* stocks
EW	20.00	1.12	1.01	0.99
	50.00	1.22	1.07	1.02
	100.00	1.26	1.09	1.03
	200.00	1.29	1.11	1.04
CW	20.00	1.02	0.95	0.95
	50.00	1.03	0.97	0.95
	100.00	1.03	1.00	0.96
	200.00	1.03	0.98	0.93

Table 9: Biases of tracking error forecasts estimated between January 2015 and December 2022 with reactive covariance matrices.

What appears from the above calculations is that the models calibrated on funds are in fact fit for the optimisation of stock portfolios. Their forecasting performance is similar to models calibrated on stocks, which is itself comparable to normalized sample covariance matrices. Also, we have measured the variance of the biases. A high variance would suggest that the performance of the model, in terms of forecastability, is indeed quite unstable over time. We found that the variance of the biases is almost identical between the different model, although de sample covariance matrix tends to exhibits somewhat higher values, thus suggesting an overall lower stability.

To sum up, the above results suggest that the same models can

be used to both analyze portfolios of funds, and optimize individual funds at the equity level. The same risk factors are used to estimate the characteristics of both funds and stocks, which insures consistency between the risk assessments made at the global and portfolio level.

12 *Conclusion*

In this paper, we present an implementation of the IPCA risk model that uses exposures to risk factors as characteristics. Among other advantages, we have shown that this type of characteristics produces risk models with excellent performances. This new implementation is particularly useful in the context of portfolio of funds, for which current models are unpractical due to the lack of holdings data.

We have shown that the risk models thus obtained exhibited high precision with a high level of consistency across instruments and time. We have also proposed solutions to the problem of find the right calibration set for these models, as these challenges were not present in the previously studied context of equities.

Also, we have discussed the stability of this model compared to traditional PCA and cross-sectional models. We have found that the dimensionality reduction mechanism embeded in the IPCA model made it particularly stable and amenable to portfolio optimisation task.

Finally, we present a number of analytical formulas that enable the utilisation of the model for high-granularity risk and performance analysis. These analytics were not provided in the literature, in spite of their importance for practical use. We have illustrated the potential of these analytics on a sample of US mutual funds.

References

Dong Hyun Ahn, Jennifer Conrad, and Robert F Dittmar. Basis assets. *Review of Financial Studies*, 22, 2009. ISSN 08939454. DOI: 10.1093/rfs/hhp065.

Noël Amenc and Felix Goltz. Long-term rewarded equity factors: What can investors learn from academic research? *The Journal of Index Investing*, 7:39–56, 2016. ISSN 2154-7238. DOI: 10.3905/jii.2016.7.2.039.

Dante Amengual and Mark W. Watson. Consistent estimation of the number of dynamic factors in a large n and t panel. *Journal of Business and Economic Statistics*, 25, 2007. ISSN 07350015. DOI: 10.1198/073500106000000585.

Jushan Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71:135–171, 2003. ISSN 00129682. DOI: 10.1111/1468-0262.00392.

Jushan Bai and Serena Ng. Large dimensional factor analysis, 2009. ISSN 15513076.

Xi Bai, Katya Scheinberg, and Reha Tutuncu. Least-squares approach to risk parity in portfolio selection. *Quantitative Finance*, 16:357–376, 2016. ISSN 14697696. DOI: 10.1080/14697688.2015.1031815.

David A. Belsley. A guide to using the collinearity diagnostics. *Computer Science in Economics and Management*, 4, 1991. ISSN 09212736. DOI: 10.1007/BF00426854.

Svetlana Bryzgalova, Markus Pelger, and Jason Zhu. Forest through the trees: Building cross-sections of stock returns. *SSRN pre print*, 2019. DOI: 10.2139/ssrn.3493458.

Edwin Burmeister, Stephen A. Ross, and N. Kahn. A practitioner's guide to factor models. *CFA Institute*, 1994.

John H Cochrane. *Asset pricing: (Revised Edition)*. Princeton University Press, 2009.

Kent Daniel and Sheridan Titman. Characteristics or covariances? *The Journal of Portfolio Management*, 24, 1998. ISSN 0095-4918. DOI: 10.3905/jpm.1998.24.

Kent Daniel and Sheridan Titman. Evidence on the characteristics of cross sectional variation in stock returns. *Journal of Finance*, 52:1–33, 7 2005. ISSN 00221082. DOI: 10.2307/2329554.

James L. Davis, Eugene F. Fama, and Kenneth R. French. Characteristics, covariances, and average returns: 1929 to 1997. *Journal of Finance*, 55, 2000. ISSN 00221082. DOI: 10.1111/0022-1082.00209.

Eugene F Fama. Eugene f. fama - prize lecture: Two pillars of asset pricing. 2013.

Eugene F. Fama and Kenneth R. French. The cross-section of expected stock returns. *Journal of Finance*, 47:427–465, 6 1992. ISSN 00221082. DOI: 10.2307/2329112.

Eugene F. Fama and Kenneth R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116:1–22, 4 2015. ISSN 0304405X. DOI: 10.1016/j.jfineco.2014.10.010.

Eugene F. Fama and Kenneth R. French. Comparing cross-section and time-series factor models. *Review of Financial Studies*, 33:1891–1926, 5 2020. ISSN 14657368. DOI: 10.1093/rfs/hhz089.

Jianqing Fan, Yuan Liao, and Weichen Wang. Projected principal component analysis in factor models. *Annals of Statistics*, 44:219–254, 2 2016. ISSN 00905364. DOI: 10.1214/15-AOS1364.

Joachim Freyberger, Andreas Neuhierl, and Michael Weber. Dissecting characteristics nonparametrically. *Review of Financial Studies*, 33, 2020. ISSN 14657368. DOI: 10.1093/rfs/hhz123.

Andrew Frongello. Linking single period attribution results. *Journal of Performance Measurement*, 6, 2002.

Michael R. Gibbons, Stephen A. Ross, and Jay Shanken. A test of the efficiency of a given portfolio. *Econometrica*, 57:1121, 9 1989. ISSN 00129682. DOI: 10.2307/1913625.

D Goldfarb and G Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28:1–38, 6 2003. ISSN 0364765X. DOI: 10.1287/moor.28.1.1.14260.

Lars Peter Hansen and Ravi Jagannathan. Assessing specification errors in stochastic discount factor models. *Journal of Finance*, 52, 1997. ISSN 00221082. DOI: 10.1111/j.1540-6261.1997.tb04813.x.

Bruce I. Jacobs and Kenneth N. Levy. Factor modeling: The benefits of disentangling cross-sectionally for explaining stock returns. *The Journal of Portfolio Management*, 47:33–50, 4 2021. ISSN 0095-4918. DOI: 10.3905/JPM.2021.1.240. URL `https://jpm.pm-research.com/content/early/2021/04/07/jpm.2021.1.240https://jpm.pm-research.com/content/early/2021/04/07/jpm.2021.1.240.abstract`.

Yindeng Jiang and Joseph Saenz. Comparing performance attribution linking methods: An empirical study. *SSRN pre print*, 6 2014. DOI: 10.2139/ssrn.2463378.

Bryan T. Kelly, Seth Pruitt, and Yinan Su. Some characteristics are risk exposures, and the rest are irrelevant. *SSRN Electronic Journal*, 9 2017. DOI: 10.2139/ssrn.3032013.

Bryan T. Kelly, Seth Pruitt, and Yinan Su. Instrumented principal component analysis. *SSRN Electronic Journal*, 4 2018. DOI: 10.2139/ssrn.2983919.

Bryan T. Kelly, Seth Pruitt, and Yinan Su. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134:501–524, 12 2019. ISSN 0304405X. DOI: 10.1016/j.jfineco.2019.05.001.

Olivier Ledoit and Michael Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10, 2003. ISSN 09275398. DOI: 10.1016/S0927-5398(03)00007-0.

Jose Menchero and Jun Wang. The barra us equity model (use4), 2011.

Kaare Brandt Petersen and Michael Syskind Pedersen. *The Matrix Cookbook*. Technical University of Denmark, 2012.

Stephen A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13:341–360, 1976. ISSN 10957235. DOI: 10.1016/0022-0531(76)90046-6.

William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk*. *The Journal of Finance*, 19:425–442, 9 1964. ISSN 00221082. DOI: 10.1111/j.1540-6261.1964.tb02865.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.1964.tb02865.x.

James H. Stock and Mark W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97:1167–1179, 12 2002. DOI: 10.2307/3085839. URL https://www.jstor.org/stable/3085839.

A Construction fundamental risk factors

Style factors

The fundamental risk factors used throughout this papers are provided by Scientific Portfolio. Figure 17 provides a summary of the characteristics used to build each of these risk factor, which we refer to as *rewarded risk factors*. The market factor is simply a market-cap weighted portfolio of the 500 largest stocks present in the US market.

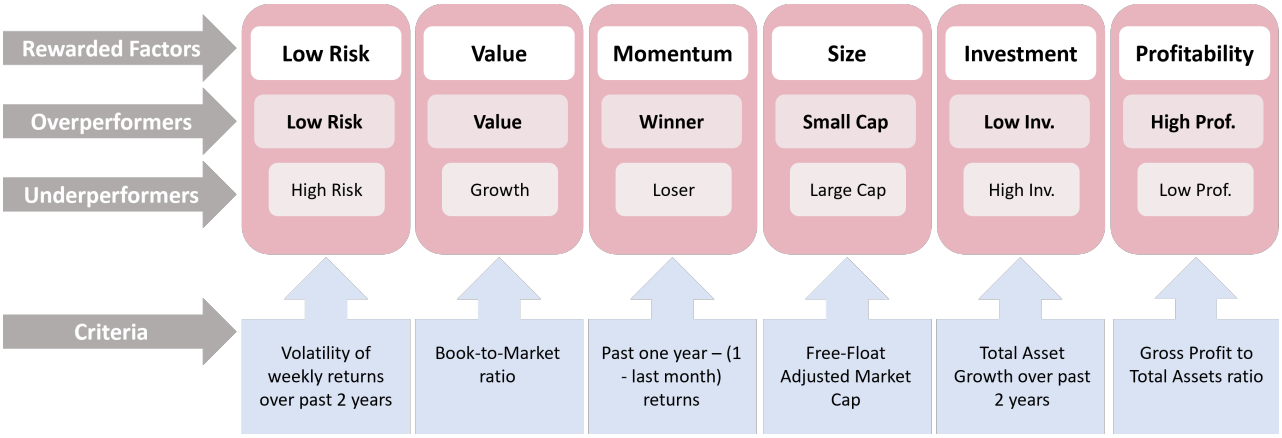

Figure 17: Summary of the construction of ScientificBeta rewarded factors.

While the market factor is cap-weighted, stocks in the long and short leg of each regressor are equally-weighted. Each regressor (with the exception of the market factor) is constructed by longing the 20% of stocks with the highest performance expectations with respect to a given characteristics (overperformers) and shorting the 20% of stocks with the lowest performance expectation (underperformer). For instance, the long leg of the "Size" factor contains the 20% smallest companies with equal weights.

The relative weight of each leg in the final portfolio is calculated so as to cancel the exposure of the factor to the market factor. At the end of each quarter, the market exposure is computed for both the long and short leg using daily returns. The relative weight of each leg is then adjusted as follows so as to neutralize the market exposure of the factor:

$$\boldsymbol{F}_{\ell,t} = \frac{(F^{\text{long}}_{\ell,t} - r_{\text{RF,t}})}{\beta^{\text{long}}_{\ell,t}} - \frac{(F^{\text{short}}_{\ell,t} - r_{\text{RF,t}})}{\beta^{\text{short}}_{\ell,t}}$$

where e.g. $\beta^{\text{long}}_{\ell,t}$ is the market beta of the long factor and $r_{\text{RF,t}}$ is the risk-free rate.

Industry factors

We use the TRBC classification standard for the regional sectors for our definition of industrial sectors. Industry risk factors correspond to the track of a portfolio that is long the sector cap-weighted index and short the market index in such a way that the resulting time-series has no exposure to the market.

B *Derivation of the gamma matrix estimation in the alternating least square algorithm*

Once the factors f_t^{new} have been estimated, the algorithm proceeds to the estimation of the $\mathbf{\Gamma}_{new}$ matrix. The new gamma matrix corresponds to the solution of the equation:

$$\mathbf{\Gamma}_{new} = \underset{\mathbf{\Gamma}}{\operatorname{argmin}} \sum_{i,t} \underbrace{(r_{i,t} - \boldsymbol{c}_{i,t}\mathbf{\Gamma}\, \boldsymbol{f}_t)^T (r_{i,t} - \boldsymbol{c}_{i,t}\mathbf{\Gamma}\, \boldsymbol{f}_t)}_{\boldsymbol{G}_t}$$

where the value of the factors have been fixed to their new value, that is $\boldsymbol{f}_t = \boldsymbol{f}_t^{new}$. The equation we have to solve is thus

$$\frac{\partial}{\partial \mathbf{\Gamma}} \sum_{i,t} \boldsymbol{G}_t = \frac{\partial}{\partial \mathbf{\Gamma}} \sum_{i,t} \left[-2\boldsymbol{f}_t^T \mathbf{\Gamma}^T \boldsymbol{c}_{i,t}^T r_{i,t} + \boldsymbol{f}_t^T \mathbf{\Gamma}^T \boldsymbol{c}_{i,t}^T \boldsymbol{c}_{i,t} \mathbf{\Gamma} \boldsymbol{f}_t \right] = 0$$

With the help of the following relationships [Petersen and Pedersen, 2012]:

$$\begin{aligned} \frac{\partial \boldsymbol{A}^T \boldsymbol{X} \boldsymbol{B}}{\partial \boldsymbol{X}} &= \boldsymbol{B}\boldsymbol{A}^T \\ \frac{\partial \boldsymbol{a}^T \boldsymbol{X} \boldsymbol{B} \boldsymbol{X} \boldsymbol{a}}{\partial \boldsymbol{X}} &= (\boldsymbol{B}^T + \boldsymbol{B}) \boldsymbol{X} \boldsymbol{a} \boldsymbol{a}^T \\ \operatorname{vec}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}) &= (\boldsymbol{B}^T \otimes \boldsymbol{A}) \operatorname{vec}(\boldsymbol{X}) \end{aligned}$$

we obtain an equation for the vectorized form of the matrix $\mathbf{\Gamma}_{new}$ as follows:

$$\frac{\partial}{\partial \mathbf{\Gamma}} \sum_{i,t} 2 f_t^T \mathbf{\Gamma}^T c_{i,t}^T r_{i,t} = \frac{\partial}{\partial \mathbf{\Gamma}} \sum_{i,t} f_t^T \mathbf{\Gamma}^T c_{i,t}^T c_{i,t} \mathbf{\Gamma} f_t$$

$$\sum_{i,t} c_{i,t}^T r_{i,t}^T f_t^T = \sum_t c_{i,t}^T c_{i,t} \mathbf{\Gamma}\, f_t f_t^T$$

$$\text{vec}(\sum_{i,t} c_{i,t}^T r_{i,t}^T f_t^T) = \text{vec}(\sum_t c_{i,t}^T c_{i,t} \mathbf{\Gamma}\, f_t f_t^T)$$

$$\sum_t (f_t \otimes c_t^T)\, \text{vec}(r_{i,t}) = \left[\sum_{i,t} (f_t^T f_t \otimes c_{i,t} c_{i,t}^T)\right] \text{vec}(\mathbf{\Gamma})$$

$$\text{vec}(\mathbf{\Gamma}) = \left[\sum_{i,t} (f_t^T f_t \otimes c_{i,t} c_{i,t}^T)\right]^{-1} \sum_{i,t} (f_t \otimes c_{i,t}^T)\, \text{vec}(r_{i,t})$$

C *Euler decomposition*

Before proceeding to the derivation of the Euler formula that is used to decompose risk and rewards, it is important to note that both of these functions not only share parameters, but also an important property: they are sensitive to leverage. Leverage sensitivity means that when portfolio weights are leveraged (multiplied) by a factor α, the risk and returns also scale by a factor proportional to α. This is expected as increasing leverage amplifies returns but also increases risk.

The fact that the functions measuring risk and rewards are sensitive to leverage is not only desirable from a financial point of view but also results in making the function "decomposable". This surprising fact is a consequence of the well-known Euler theorem that states whenever a function scales with its parameter (here denoted by $\boldsymbol{x}$), that is:

$$f(\alpha \boldsymbol{x}) = \alpha^{\nu} f(\boldsymbol{x}),$$

it can be decomposed into contributions attributable to the parameters. Here, the variable ν can be any number and impacts the proportionality between the value of the scaling parameter α and response from the function. Any function with the above mathematical property is called *homogeneous*.

In our case, we can use both the portfolio's weights or the characteristics as parameters. For ease of notation, we group the instrument characteristics and specific risks into one variable $\boldsymbol{y}$. For portfolios, the characteristics are transformed into weighted characteristics. Although they are two different parameters, we do not need to separate

characteristics from weights. This is because scaling the characteristics has the same effect as scaling the weights: we can scale the characteristics while maintaining $\boldsymbol{w}$ fixed, and vice-versa to obtain the same effect. Because we are interested in understanding how exposures to fundamental factors affect risk, we focus on characteristics and hence use $\boldsymbol{y}$ as a parameter. Note that for both risk and return functions, $\nu = 1$, which means that the leverage actually scales linearly.

The Euler decomposition is based on the following observation. First recall the definition of a function's total derivative:

$$\mathrm{d}f(\boldsymbol{y}) = \sum_i \frac{\partial f(\boldsymbol{y})}{\partial y_i} \mathrm{d}y_i$$

The Euler formula is obtained by introducing a variable $\boldsymbol{x}$ so that $\boldsymbol{y} = \alpha \boldsymbol{x}$ and $d\boldsymbol{y}/\mathrm{d}\alpha = \boldsymbol{x}$. Using the homogeneity of the function $f(\alpha \boldsymbol{x})$:

$$\frac{\mathrm{d}f(\alpha \boldsymbol{x})}{\mathrm{d}\alpha} = \nu \alpha^{\nu-1} f(\boldsymbol{x}) = \sum_i \frac{\partial f(\boldsymbol{y})}{\partial y_i} \frac{dy_i}{\mathrm{d}\alpha}$$

If we cancel the effect of the leverage by setting $\alpha = 1$, then $\boldsymbol{x} = \boldsymbol{y}$ and we are left with

$$\nu f(\boldsymbol{x}) = \sum_i \frac{\partial f(\boldsymbol{x})}{\partial x_i} x_i \text{ (Euler decomposition)}$$

The interpretation of the Euler formula is very simple: if a function scales with its parameters, then it is decomposable into an addition of contributions. Each contribution is composed of the function's first-order sensitivity (partial derivative) to a given parameter times the parameter itself. Because this decomposition is reminiscent of a linear combination where the sensitivities are the coefficients and the parameters the variables, we can interpret these as the contribution of the i^{th} parameter to the function's total value. This is why the Euler formula is used extensively in the field of finance to understand the drivers of risk and returns.

D Annualization of performance

In Sec. 11, we have only considered the total performance of instruments. In many cases, especially when the period considered spans

several years, the annualized performance is preferred.

For weekly returns, the annualized performance is given by:

$$r_w^{\text{p.a.}} = (1 + r_w)^{\frac{1}{n}} - 1$$

where n denotes the number of years covered by the period over which r_w is estimated. To the first order, and assuming that r_w is rather small, the annualized performance is given by:

$$r_w^{\text{p.a.}} \approx \underbrace{f(0)}_{=0} + f(0)' r_w = \frac{1}{n} r_w$$

The first order thus simply corresponds to the total return divided by the number of years. It is reminiscent of the "street convention" used for yield-to-maturity of short-term bonds. To the second order, the above approximation becomes:

$$\begin{aligned} r_w^{\text{p.a.}} &\approx f(0)' r_w + \frac{1}{2} f(0)'' r_w^2 \\ &\approx \frac{1}{n} r_w + \frac{1}{n} \frac{1-n}{2n} r_w^2 \\ &\approx \frac{1}{n} r_w \left(1 + \frac{1-n}{2n} r_w\right) \end{aligned}$$

Grouping terms, we find that the total return is approximately annualized by the following adjustment:

$$\begin{aligned} r_w^{\text{p.a.}} &\approx \eta \, r_w \text{ with} \\ \eta &= \frac{1}{n} \left(1 + \frac{1-n}{2n} r_w\right) \end{aligned}$$

In fact, it is easy to show that developing the above formula for higher orders yields:

$$\begin{aligned} r_w^{\text{p.a.}} &= \sum_{j=1}^{\infty} \frac{f^{(j)}(0)}{n!} r_w^j = \sum_{j=1}^{\infty} \binom{1/n}{j} r_w^j = \eta \, r_w \\ \text{with } \eta &= \sum_{j=1}^{\infty} \binom{1/n}{j} r_w^{j-1} \end{aligned}$$

Here $\binom{\alpha}{k} = \prod_{k=1}^{n}(\alpha - k + 1)/k$ denotes the generalized binomial coefficient. Consequently, multiplying the returns by a well chosen factor is an efficient way to annualize returns. This factor is simply obtained by taking:

$$\eta = \frac{r_w}{r_w^{\text{p.a.}}}$$

The advantage of such a factor is that it can be applied to all contributions to performance to produce annualized contributions. Combined with the Frongello adjustment factor, it produces contributions that are both summable and annualized.

E Procruste's uniquely identifiable class of parameters

As we have explained in Sect. 9, the convergence properties of the calibration algorithm requires the parametes to be part of a uniquely identifiable class. Uniquely identifiable parameters have pre-defined properties that constrain their values.

Beyond the numerical convergence of the calibration algorithm, having parameters with a number of predefined properties brings a number of advantages. Notably, imposing the orthogonality of the $\boldsymbol{f_t}$ factors enables to simplify greatly a number of calculations involving the model. In the original algorithm proposed in [Kelly et al., 2017], the factors are set to be orthogonal and the lines of the Γ matrix pairwise orthonormal. Another uniquely identifiable class proposed by the authors involved forcing certain elements of the gamma matrix to one in order to facilitate the financial interpretation of the factors.

When using the model to derive analytical relationships, we found that only the orthogonality of the factors $\boldsymbol{f_t}$ was really useful, whereas the properties of the gamma matrix did not matter much. However, in some cases, imposing some structure to the gamma matrix has some practical interest. Specifically, imposing to the gamma matrix to be as close as possible to a reference, while preserving the factors orthogonality, has a number of applications. For instance, it is applicable to the stabilisation of the gamma matrix when the estimation period is changed, or when small changes in the calibration inputs are made. In this case the reference matrix will be set to the previously calculated gamma. Also, it maybe useful for testing purpose to require the market beta to be associated with only one statistical factor.

The following transformation allows to impose the factors to be

orthogonal while maintaining a high level of similarity between the optimal gamma matrix and a reference matrix. First, let us recall that the model parameters are rotationally invariant, which means there exist a transformation $\boldsymbol{R}$ such that

$$\begin{aligned}\boldsymbol{\Gamma}' &= \boldsymbol{\Gamma}\,\boldsymbol{R}\\ \boldsymbol{f}'_t &= \boldsymbol{R}^{-1}\boldsymbol{f}_t\end{aligned}$$

where R is a $k \times k$ matrix. As we have explained in Sect. 9, the factors orthogonality was obtained by setting $\boldsymbol{R}$ to the unitary[12] transformation corresponding to the left-most element of the factors SVD decomposition. In the sequel, we will denote this transformation by $\boldsymbol{U}$.

[12] The transpose of a unitary matrix is also its inverse, so that $\boldsymbol{U}^T\boldsymbol{U} = 1$.

If we write the transformation $\boldsymbol{R}$ as a multiplication of $\boldsymbol{U}$ by another unitary transformation $\boldsymbol{A}$, that is

$$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{A},$$

then $\boldsymbol{R}$ still orthogonalizes the factors. So the presence of $\boldsymbol{A}$ in the definition of $\boldsymbol{R}$ does not affect the orthogonality of the factors. The question is now how to define $\boldsymbol{A}$. One way of doing so is to use $\boldsymbol{A}$ so as to make the gamma matrix as close as possible to a reference, or in mathematical terms

$$\boldsymbol{A} = \arg\min_{A} ||\boldsymbol{\Gamma}_{ref} - \boldsymbol{\Gamma}\boldsymbol{U}\boldsymbol{A}||_F \text{ with } \boldsymbol{A}^T\boldsymbol{A} = \mathbf{1}$$

where $||\circ||_F$ denotes the Frobenius[13] norm. This formulation is reminiscent of the *Procrustes problem* and is equivalent to

[13] The Frobenius norm is defined as $||\boldsymbol{A}||_F = Tr[\boldsymbol{A}\boldsymbol{A}^T]$.

$$\boldsymbol{A} = \arg\min_{A} Tr[\boldsymbol{\Gamma}_{ref}^T\boldsymbol{\Gamma}\boldsymbol{U}\boldsymbol{A}]$$

for which a solution is readily available.