NBER WORKING PAPER SERIES

WHEN DO CROSS-SECTIONAL ASSET PRICING FACTORS SPAN THE STOCHASTIC DISCOUNT FACTOR?

Serhiy Kozak
Stefan Nagel

Working Paper 31275
http://www.nber.org/papers/w31275

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2023

We thank Mike Chernov, Kent Daniel, Gene Fama, Stefano Giglio, Lars Hansen, Ian Martin, Andreas Neuhierl, Markus Pelger, seminar participants at City University of Hong Kong, Stanford University, Johns Hopkins University, Swiss Finance Institute in Lugano, Tulane University, University of North Carolina at Chapel Hill, University of Maryland, and conference participants at the AI & Big Data Research Forum, American Finance Association meetings, SoFiE, UCLA Backus Memorial Conference for helpful comments and discussions. We also thank Tianshu Lyu for excellent research assistance. Nagel gratefully acknowledges financial support from the Fama-Miller Center and the Center for Research in Security Prices at the University of Chicago Booth School of Business. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

When Do Cross-Sectional Asset Pricing Factors Span the Stochastic Discount Factor?
Serhiy Kozak and Stefan Nagel
NBER Working Paper No. 31275
May 2023
JEL No. G11,G12

ABSTRACT

When expected returns are linear in asset characteristics, the stochastic discount factor (SDF) that prices individual stocks can be represented as a factor model with GLS cross-sectional regression slope factors. Factors constructed heuristically by aggregating individual stocks into characteristics-based factor portfolios using sorting, characteristics-weighting, or OLS cross-sectional regression slopes do not span this SDF unless the covariance matrix of stock returns has a specific structure. These conditions are more likely satisfied when researchers use large numbers of characteristics simultaneously. Methods to hedge unpriced components of heuristic factor returns allow partial relaxation of these conditions. We also show the conditions that must hold for dimension reduction to a number of factors smaller than the number of characteristics to be possible without having to invert a large covariance matrix. Under these conditions, instrumented and projected principal components analysis methods can be implemented as simple PCA on characteristics-based portfolios.

Serhiy Kozak
University of Maryland
7621 Mowatt Ln
VMH 4453
College Park, MD 20742
sekozak@umd.edu

Stefan Nagel
University of Chicago
Booth School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
stefan.nagel@chicagobooth.edu

I. Introduction

Cross-sectional asset pricing research has linked stocks' expected returns to a large set of firm characteristics. To summarize these cross-sectional pricing patterns in a reduced-form pricing model, researchers often construct stochastic discount factor (SDF) proxies with multiple characteristics-based factors. Individual assets' weights in these factor portfolios are functions of stock characteristics. Researchers use a variety of different heuristic approaches to specify these weights. For example, Fama and French (1993) sort stocks by characteristics and then form portfolios by applying quintile cutoffs (*sorted factors*); Kozak, Nagel, and Santosh (2020) construct portfolios with weights proportional to stocks' centered univariate cross-sectional rank for each characteristic (*univariate factors*); Fama and French (2020) use the slopes of monthly cross-sectional OLS regressions of returns on characteristics as factor portfolio returns (*OLS factors*).[1]

There is, however, only one unique SDF—or, equivalently, one mean-variance efficient portfolio return—that is spanned by returns of the individual assets (Hansen and Jagannathan 1991). Under which conditions do these different heuristic methods yield this SDF? Put differently, under which conditions is the investment opportunity set not deteriorating if one aggregates individual assets to these factor portfolios? Somewhat surprisingly, the answer to this fundamental question is not available in the literature.

Clearly, some special conditions must be met because the weights of individual assets in the mean-variance efficient portfolio weights depend on the return covariance matrix, but none of these heuristic methods use any information from the covariance matrix in factor construction. Our first objective in this paper is to work out what these conditions are.

We set estimation issues aside at first and work with population moments. We assume

1. More precisely, Fama and French (2020) use a hybrid approach where individual stocks are first sorted into a relatively large number of characteristics-based portfolios and the value-weighted returns and average characteristics of these portfolios are then the input for the cross-sectional OLS regressions. The MSCI BARRA factor model, which goes back to Rosenberg (1974) and is widely applied in industry, is also based on OLS factors.

that conditional expected returns of N individual stocks are linear in $J \leq N$ firm characteristics collected in the $N \times J$ characteristics matrix $\boldsymbol{X}_t$. At this conceptual level, this assumption is very general since the set of characteristics could also contain nonlinear functions of some underlying basic characteristics. For instance, sorted factors can be subsumed by univariate factors if characteristics are specified as step functions. This linearity assumption only acquires substantive empirical content once a researcher has fixed a specific set of characteristics that she works with.

As a starting point, we show that the SDF that prices all individual assets can be expressed as a multifactor SDF spanned by J factors that are the slopes of cross-sectional GLS regressions of returns on lagged firm characteristics (*GLS factors*). The inverse of the conditional covariance matrix of returns serves as the GLS weighting matrix. The matrix of individual assets' conditional betas on the GLS factors is then exactly equal to $\boldsymbol{X}_t$.[2]

In practice, construction of these GLS factors would be difficult because it requires estimating and inverting a large conditional covariance matrix. For this reason, it is important to know whether heuristic approaches that bypass this inversion problem can deliver factors that span the SDF. Sorted factors, univariate factors, and OLS factors are all simply weighting stocks by columns of $\boldsymbol{X}_t$ or a nonsingular linear transformation thereof. We show that these factors span the SDF if and only if the conditional covariance matrix $\boldsymbol{\Sigma}_t$ of individual asset returns takes the specific form

$$\boldsymbol{\Sigma}_t = \boldsymbol{X}_t \boldsymbol{\Psi}_t \boldsymbol{X}_t' + \boldsymbol{U}_t \boldsymbol{\Omega}_t \boldsymbol{U}_t', \qquad \boldsymbol{X}_t' \boldsymbol{U}_t = \boldsymbol{0}. \tag{1}$$

This means that there must be a clean separation among the sources of systematic risk such that loadings on up to J systematic factors are perfectly spanned by $\boldsymbol{X}_t$ while loadings on the remaining ones are orthogonal to $\boldsymbol{X}_t$. When (1) holds, individual assets' betas on OLS factors are exactly equal to $\boldsymbol{X}_t$, i.e., covariances are equal to characteristics. Fama and

2. The GLS factors are similar to the characteristics-efficient portfolios of Daniel, Mota, Rottke, and Santos (2020), but in our analysis we allow for time-varying conditional moments.

French (2020) argue that the OLS factors can be used as asset pricing factors in time-series regressions with conditional betas set equal to $\boldsymbol{X}_t$, but our result shows that this is true if and only if condition (1) is satisfied.

Condition (1) is more likely to hold approximately when $\boldsymbol{X}_t$ includes a large, comprehensive set of characteristics. In this case, important sources of stock return covariance can be absorbed in the first term of $\boldsymbol{\Sigma}_t$ in (1), which leaves $\boldsymbol{U}_t$ and violations of $\boldsymbol{X}_t'\boldsymbol{U}_t = \boldsymbol{0}$ quantitatively unimportant. Additional characteristics can help even if they are unrelated to expected returns as long as they help to capture major sources of stock return covariances. But if the number of characteristics is small—as in popular low-dimensional factor models with only four or five characteristics-based factors—there is little reason to think that this small number of characteristics should be sufficient to span loadings on all major sources of covariance.

Importantly, the mean-variance inefficiency of heuristic factor models that we study does not come from omitting characteristics. If condition (1) does not hold, heuristic factors are inefficient relative to GLS factors even though they use the same characteristics. The inefficiency of heuristic factors comes from contamination with unpriced risks that are not compensated with higher mean returns. This is different from, e.g. Giglio and Xiu (2021), who study biases in factor risk premia estimates due to omitted priced factors which are likely associated with omitted characteristics.

Existing empirical results in the literature show that commonly-used heuristic factors are contaminated with unpriced risks, which suggests that low-dimensional models based on these factors do not satisfy condition (1).[3] Motivated by these findings, researchers have developed heuristic methods to remove unpriced components from heuristic factors. Daniel, Mota, Rottke, and Santos (2020) (DMRS) construct hedge portfolios that have positive loadings

3. For example, Gerakos and Linnainmaa (2018) find that the HML value factor is contaminated with unpriced components; Back, Kapadia, and Ostdiek (2015) find that OLS factors have alpha with respect to the standard sorted factors of Hou, Xue, and Zhang (2015) and Fama and French (2015); Grinblatt and Saxena (2018) find that sorted factors do not price the basis portfolios from which they were constructed; Chib, Lin, Pukthuanthong, and Zeng (2021) find that the method of factor construction affects asset pricing performance.

on the original factors but zero exposure to the underlying characteristics that determine expected returns. Residualizing the original factors with respect to the hedge portfolio returns removes unpriced risks. However, it is not clear under which conditions this heuristic hedging approach actually yields a better approximation of the SDF. Our second objective therefore is to understand the conditions under which this hedging approach can be used to recover factors that span the SDF.

We show that the hedged factors span the SDF if the covariance matrix has the structure in (1), but with the requirement $\boldsymbol{X}_t'\boldsymbol{U}_t$ replaced with the requirement that there exists a decomposition such that

$$\boldsymbol{U}_t\boldsymbol{\Omega}_t\boldsymbol{U}_t' = \boldsymbol{V}_t\boldsymbol{\Gamma}_t\boldsymbol{V}_t' + \boldsymbol{E}_t\boldsymbol{\Phi}_t\boldsymbol{E}_t', \qquad \boldsymbol{X}_t'\boldsymbol{E}_t = \boldsymbol{0}, \tag{2}$$

where $\boldsymbol{V}_t$ is an $N \times J$ matrix. This is a weaker condition than $\boldsymbol{X}_t'\boldsymbol{U}_t$ because here columns of $\boldsymbol{X}_t$ can be correlated with columns of $\boldsymbol{U}_t$, as long as this correlation comes only through the J columns of $\boldsymbol{V}_t$. Again, this condition is more likely to hold when researchers consider a large, comprehensive set of characteristics.

While DMRS consider only one round of hedging, there is no reason to stop after one round. We show that iteration on this approach, by hedging once more the already-hedged factor portfolios can yield further improvements and further weakens the requirements on the covariance matrix. One can think of this iterated hedging approach as incorporating more and more information from the covariance matrix into the factors which brings them closer to GLS factors.

The approaches we discussed so far construct J factors to capture the pricing information of J characteristics. Dimension-reduction methods aim to span the SDF with a smaller number of $K < J$ factors while again avoiding the need to invert an estimate of $\boldsymbol{\Sigma}_t$. Different approaches for dimension reduction exist in the literature, but it is not clear what the necessary conditions are for the factors constructed with these methods to span the SDF. Our third objective is therefore to establish these conditions.

We show that if and only if the conditional covariance matrix has a structure like in (1), but with $\boldsymbol{X}_t$ replaced by lower-dimensional $K \leq J$ linear combinations of characteristics collected in $\boldsymbol{X}_t\boldsymbol{Q}_t$, then portfolios with weights equal to $\boldsymbol{X}_t\boldsymbol{Q}_t$, or a non-singular linear transformation thereof, span the SDF. We further show that under these conditions, two prominent methods of dimension reduction, the instrumented principal components method (IPCA) of Kelly, Pruitt, and Su (2019) and the projected PCA method (PPCA) of Kim, Korajczyk, and Neuhierl (2021) are closely related. IPCA can then be implemented using simple PCA on OLS factors while PPCA can be implemented via simple PCA on univariate factors constructed using orthonormalized characteristics.

Finally, we turn to empirical implementation. Our theoretical results are all stated in terms of population moments. However, we find that our theoretical results also characterize well the properties of factor models with empirically estimated moments.

In the first part of our empirical analysis, we primarily focus on the properties of OLS factor models constructed using the stock characteristics from Kozak (2019) and Kozak, Nagel, and Santosh (2020). Consistent with our theoretical results, OLS factors generally do not span the SDF that prices individual stocks. We infer this from the fact that hedging the OLS factors' unpriced risk exposures, or constructing approximate GLS factors, produces statistically significant improvements of the maximum squared Sharpe ratio attainable in- and out-of-sample. Iterating the hedging procedures produces further gains in the maximum squared Sharpe ratio. Furthermore, while these gains are large for small-scale factor models that use only a few characteristics, they vanish when we use a large number of characteristics to construct the OLS factors. This is in line with our conclusion from the theoretical analysis that condition (1) is more likely to hold, and therefore OLS factors more likely to span the SDF that prices individual stocks, when the econometrician employs a large number of characteristics. Interestingly, improvements from hedging are bigger for univariate factors than for OLS factors, which suggests that univariate factors are more contaminated by unpriced risks.

In the second part of our empirical analysis we implement and test several methods of dimensionality reduction based on Kozak, Nagel, and Santosh (2020), Kelly, Pruitt, and Su (2019), and Kim, Korajczyk, and Neuhierl (2021). We find that latent factor models perform quite differently depending on how their factors are constructed. As in the case without dimension reduction, applying the OLS transformation to characteristics yields more efficient factors that are less contaminated with unpriced risks.

II. Conditions for characteristics-based portfolios to span the mean-variance frontier

We consider a cross-section of N assets with an $N \times 1$ vector of excess returns $\boldsymbol{z}_{t+1}$. Each asset features J characteristics that are observable to the econometrician, collected in the (time-varying) $N \times J$ matrix $\boldsymbol{X}_t$ where $J \leq N$, $\text{rank}(\boldsymbol{X}_t) = J$, and the first column of $\boldsymbol{X}_t$ is a vector of ones. In a number of places in our analysis we will use the residual maker matrix $\boldsymbol{R}_t = \boldsymbol{I} - \boldsymbol{X}_t \left(\boldsymbol{X}_t'\boldsymbol{X}_t\right)^{-1} \boldsymbol{X}_t'$ that generates the residuals in a projection on $\boldsymbol{X}_t$. Unless otherwise noted, we use the notation $\boldsymbol{\mu}_{y,t} = \mathbb{E}_t[\boldsymbol{y}_{t+1}]$, $\boldsymbol{\Sigma}_{y,t} = \text{var}_t(\boldsymbol{y}_{t+1})$ for the conditional moments of a random vector $\boldsymbol{y}_{t+1}$, $\boldsymbol{\Sigma}_{xy,t}$ as notation for the conditional covariance matrix of two random vectors $\boldsymbol{x}_{t+1}$ and $\boldsymbol{y}_{t+1}$, and $\boldsymbol{I}_K$ for a $K \times K$ identity matrix.

In what follows, all time-t conditional moments are conditioned on $\boldsymbol{X}_t$. We denote

$$\boldsymbol{\Sigma}_t = \text{var}(\boldsymbol{z}_{t+1}|\boldsymbol{X}_t), \qquad \boldsymbol{\mu}_t = \mathbb{E}[\boldsymbol{z}_{t+1}|\boldsymbol{X}_t], \tag{3}$$

and we assume that $\boldsymbol{\Sigma}_t$ is positive definite. That these conditional moments are conditioned on the characteristics observable to the econometrician is important. The set of characteristics observable to investors could be larger or smaller than what is contained in $\boldsymbol{X}_t$, without consequences for our results, as long as the law of one price holds conditional on $\boldsymbol{X}_t$.[4] There-

4. As an example that would violate this requirement, the law of one price would fail if the econometrician included elements of $\boldsymbol{z}_{t+1}$ in $\boldsymbol{X}_t$. Conditional on this look-ahead information, arbitrage opportunities would seemingly exist.

fore, it is possible that conditional on investors' information set, moments of excess returns could vary more or less than conditional on the econometrician's information. Only sources of variation linked to $\boldsymbol{X}_t$ matter in our analysis.

We assume throughout that the law of one price holds and hence an SDF exists. Conditional on the econometrician's information, the maximum squared conditional Sharpe ratio that can be obtained from the N individual assets then is finite and given by $\boldsymbol{\mu}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t$. The SDF that uses this maximum squared conditional Sharpe ratio portfolio as risk factor,

$$M_{t+1} = 1 - \boldsymbol{\delta}_t' \left(\boldsymbol{z}_{t+1} - \boldsymbol{\mu}_t \right), \qquad \boldsymbol{\delta}_t = \boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t, \tag{4}$$

prices the N assets conditionally, i.e., $\mathbb{E}[M_{t+1}\boldsymbol{z}_{t+1}|\boldsymbol{X}_t] = 0$. This is the unique SDF (with mean of unity) in the span of excess returns. We refer to it from now on simply as the SDF.

Our analysis focuses on characteristics-based factors. These factors are generally constructed with an $N \times J$ portfolio weight matrix $\boldsymbol{W}_t$, where the weights are functions of the characteristics $\boldsymbol{X}_t$, and possibly also of $\boldsymbol{\Sigma}_t$. Using these weights, one can form J factor portfolios as

$$\boldsymbol{f}_{t+1} = \boldsymbol{W}_t'\boldsymbol{z}_{t+1}, \tag{5}$$

with $\boldsymbol{\mu}_{f,t} = \boldsymbol{W}_t'\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_{f,t} = \boldsymbol{W}_t'\boldsymbol{\Sigma}_t\boldsymbol{W}_t$. We assume that weights are such that $\boldsymbol{\Sigma}_{f,t}$ is positive definite.

Our aim is to understand under which conditions different specifications of the weights $\boldsymbol{W}_t$ produce factors that span the conditional mean-variance frontier. Spanning the conditional mean-variance frontier is equivalent to the factors' maximum squared conditional Sharpe ratio,

$$\boldsymbol{\mu}_{f,t}'\boldsymbol{\Sigma}_{f,t}^{-1}\boldsymbol{\mu}_{f,t} = \boldsymbol{\mu}_t'\boldsymbol{W}_t(\boldsymbol{W}_t'\boldsymbol{\Sigma}_t\boldsymbol{W}_t)^{-1}\boldsymbol{W}_t'\boldsymbol{\mu}_t, \tag{6}$$

attaining the maximum squared conditional Sharpe Ratio obtainable from the individual assets. Our results below rely on the following lemma that provides conditions under which this is true.

Lemma 1 *The maximum squared conditional Sharpe ratio of the factors* $\boldsymbol{f}_{t+1} = \boldsymbol{W}_t' \boldsymbol{z}_{t+1}$ *is equal to the maximum squared conditional Sharpe Ratio of the individual assets, i.e.,*

$$\boldsymbol{\mu}_t' \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\mu}_t = \boldsymbol{\mu}_t' \boldsymbol{W}_t \left(\boldsymbol{W}_t' \boldsymbol{\Sigma}_t \boldsymbol{W}_t \right)^{-1} \boldsymbol{W}_t' \boldsymbol{\mu}_t \tag{7}$$

if and only if

$$\boldsymbol{\mu}_t = \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{b}_t \tag{8}$$

for some $J \times 1$ *vector* $\boldsymbol{b}_t$.

Proof. Following the proof of Lu and Schmidt (2012) Theorem 3 (A, B), express the difference of the left- and right-hand-sides of (7) as $\boldsymbol{\Delta} = \boldsymbol{\mu}_t' \boldsymbol{\Sigma}_t^{-\frac{1}{2}} \boldsymbol{M} \boldsymbol{\Sigma}_t^{-\frac{1}{2}} \boldsymbol{\mu}_t$, where $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{P}$ and $\boldsymbol{P}$ are residual and projection matrices, respectively, for a projection onto the columns of $\boldsymbol{\Sigma}_t^{\frac{1}{2}} \boldsymbol{W}_t$. $\boldsymbol{\Delta} = \boldsymbol{0}$ if and only if $\boldsymbol{\Sigma}_t^{-\frac{1}{2}} \boldsymbol{\mu}_t$ is in the column space of $\boldsymbol{\Sigma}_t^{\frac{1}{2}} \boldsymbol{W}_t$, that is, $\boldsymbol{\Sigma}_t^{-\frac{1}{2}} \boldsymbol{\mu}_t = \boldsymbol{\Sigma}_t^{\frac{1}{2}} \boldsymbol{W}_t \boldsymbol{b}_t$ for some $\boldsymbol{b}_t$, which is equivalent to (8). ■

Condition (8) has an intuitive asset pricing interpretation. It states that asset risk premia, $\boldsymbol{\mu}_t$, must come from covariances of asset returns with the factors, $\boldsymbol{\Sigma}_t \boldsymbol{W}_t$. Alternatively, if one pre-multiplies the equation with $\boldsymbol{\Sigma}_t^{-1}$, then it states that SDF risk prices or mean-variance efficient (MVE) portfolio weights, $\boldsymbol{\Sigma}_t^{-1} \boldsymbol{\mu}_t$, must be spanned by factor weights $\boldsymbol{W}_t$.

If the factors span the conditional mean-variance frontier, then they span the SDF that prices the individual assets:

Corollary 1 *Lemma 1 implies that if and only if equation* (8) *holds, an SDF can be represented in terms of the J factors:*

$$M_{t+1} = 1 - \boldsymbol{b}_t' \left(\boldsymbol{f}_{t+1} - \boldsymbol{\mu}_{f,t} \right). \tag{9}$$

This SDF perfectly prices the excess returns $\boldsymbol{z}_{t+1}$*, that is,* $\mathbb{E}\left[M_{t+1} \boldsymbol{z}_{t+1} | \boldsymbol{X}_t \right] = 0$. *This SDF representation is equivalent to a conditional beta-pricing representation*

$$\boldsymbol{\mu}_t = \boldsymbol{\beta}_t \boldsymbol{\mu}_{f,t}, \tag{10}$$

where $\boldsymbol{\beta}_t = \boldsymbol{\Sigma}_{zf,t}\boldsymbol{\Sigma}_{f,t}^{-1}$.

Equipped with this result, we can now explore under which assumptions about $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ various heuristic methods of factor construction that have appeared in the literature yield factors that span the SDF.

Our baseline assumption about expected returns is motivated by a large body of work that has documented cross-sectional relationships between expected return and firm characteristics:

Assumption 1 *(Linearity of expected returns in characteristics)*

$$\boldsymbol{\mu}_t = \boldsymbol{X}_t\boldsymbol{\phi} \tag{11}$$

for some $J \times 1$ *vector* $\boldsymbol{\phi}$.

At a conceptual level, the assumption that $\boldsymbol{\mu}_t$ is linear in $\boldsymbol{X}_t$ is without loss of generality as $\boldsymbol{X}_t$ could also include nonlinear functions of characteristics. Similarly, portfolio sorting approaches that allow expected returns to differ across but not within bins defined by characteristics can be accommodated in Assumption 1 by letting $\boldsymbol{X}_t$ be a matrix of bin membership indicators. That $\boldsymbol{\phi}$ is a constant parameter vector is not restrictive either, because one could include nonlinear interactions of cross-sectional firm characteristics with time-series predictors to capture any time-variation in expected returns. In practice, though, once a researcher has chosen a specific set of characteristics to include in $\boldsymbol{X}_t$, Assumption 1 becomes a substantive assumption that restricts $\boldsymbol{\mu}_t$. Later in the paper, we discuss alternative assumptions, including a potential modeling of MVE portfolio weights, $\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t$, as linear in $\boldsymbol{X}_t$.

Comparing Assumption 1 and equation (8), we see that $\boldsymbol{\Sigma}_t\boldsymbol{W}_t$ collapses to $\boldsymbol{X}_t$ only in special cases when certain conditions are satisfied for $\boldsymbol{W}_t$, or certain restrictions on $\boldsymbol{\Sigma}_t$ hold. We now explore these conditions.

II.A. The unique SDF in the span of excess returns: GLS factors and transformations thereof

As a benchmark for understanding when and why heuristic factor models span or do not span the SDF, we first show that the SDF in (9) has a J-factor representation under Assumption 1:

Proposition 1 *Assumption 1 is equivalent to the statement that an SDF given by* (9) *with characteristics-based factors*

$$\boldsymbol{f}_{t+1} = \boldsymbol{S}_t' \boldsymbol{X}_t' \boldsymbol{\Sigma}_t^{-1} \boldsymbol{z}_{t+1}, \tag{12}$$

and prices of risk

$$\boldsymbol{b}_t = \boldsymbol{S}_t^{-1} \boldsymbol{\phi}, \tag{13}$$

where $\boldsymbol{S}_t$ is any nonsingular $J \times J$ transformation matrix, perfectly prices the excess returns $\boldsymbol{z}_{t+1}$, that is, $\mathbb{E}\left[M_{t+1}\boldsymbol{z}_{t+1}|\boldsymbol{X}_t\right] = 0$.

Proof. Rewrite (11) as $\boldsymbol{\mu}_t = \boldsymbol{\Sigma}_t \boldsymbol{\Sigma}_t^{-1} \boldsymbol{X}_t \boldsymbol{S}_t \boldsymbol{S}_t^{-1} \boldsymbol{\phi} = \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{b}_t$, where $\boldsymbol{W}_t = \boldsymbol{\Sigma}_t^{-1} \boldsymbol{X}_t \boldsymbol{S}_t$. Lemma 1 now applies. ■

Thus, when there is a linear relationship between J characteristics and conditional expected return, the SDF is spanned by J characteristics-based factors that exactly explains these conditional expected returns with zero pricing errors. Proposition 1 therefore highlights that there is no economic difference between a model that specifies expected returns directly as linear function of characteristics as in Assumption 1 and a characteristics-based factor pricing model. One can always be mapped perfectly into the other one, with equivalent pricing implications. Therefore, a horse race between direct linear prediction of $\boldsymbol{z}_{t+1}$ by $\boldsymbol{X}_t$ and a factor pricing model, e.g., as in as in Daniel and Titman (1997) and Davis, Fama, and French (2000) as well as many other papers, does not have economic content. If factors are constructed as in Proposition 1, there is no difference in expected returns implied by direct linear prediction and the factor model. If factors are constructed in a heuristic way that does not exactly follow the prescription of Proposition 1, then there can be a difference, but this just reflects the misspecification of the heuristic factors. The difference does not have economic content (it does not discriminate between "rational" and "behavioral" asset pricing theories, for example).

Empirical asset pricing researchers often like to work with beta-pricing specifications and, in particular, with beta-pricing specifications that can be conditioned down to deliver predictions for unconditional expected returns without elaborate estimation of time-varying conditional moments. The following example present such a case.

Example 1 *Suppose* $\boldsymbol{S}_t = \left(\boldsymbol{X}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{X}_t\right)^{-1}$. *We then obtain an SDF with factors given by GLS cross-sectional regression slopes,* $\boldsymbol{f}_{t+1} = \left(\boldsymbol{X}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{X}_t\right)^{-1}\boldsymbol{X}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{z}_{t+1}$. *Factor risk prices are time varying,* $\boldsymbol{b}_t = \left(\boldsymbol{X}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{X}_t\right)\boldsymbol{\phi}$. *Factor means are constant,* $\boldsymbol{\mu}_{f,t} = \boldsymbol{\phi}$. *Factor betas are equal to characteristics,* $\boldsymbol{\beta}_t = \boldsymbol{X}_t$.

The GLS slope factors in this example are the GLS counterpart to the OLS cross-sectional slope factors in Fama (1976) and Fama and French (2020). The factors in Example 1 are also similar to the "characteristic-efficient portfolios" in Daniel, Mota, Rottke, and Santos (2020), albeit here with time-varying $\boldsymbol{X}_t$ and conditional moments of excess returns. We will show later in Section IV that keeping track of time-variation in $\boldsymbol{X}_t$ and conditional moments is important in empirical implementation of these factor models.

Which transformation matrix $\boldsymbol{S}_t$ to pick is a matter of convenience. The next example is one in which factor covariances instead of factor betas are equal to $\boldsymbol{X}_t$:

Example 2 *Suppose* $\boldsymbol{S}_t = \boldsymbol{I}$. *We then obtain "MVE factors",* $\boldsymbol{f}_{t+1} = \boldsymbol{X}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{z}_{t+1}$. *Factor risk prices are constant,* $\boldsymbol{b}_t = \boldsymbol{\phi}$. *Factor means are time-varying,* $\boldsymbol{\mu}_{f,t} = \left(\boldsymbol{X}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{X}_t\right)\boldsymbol{\phi}$. *Covariances of returns and factors are equal to characteristics,* $\boldsymbol{\Sigma}_{zf,t} = \boldsymbol{X}_t$.

Practical implementation of the SDF in Proposition 1 is of course difficult since it involves the inversion of a large $N \times N$ conditional covariance matrix. Heuristic approaches to factor construction exist that avoid this inversion problem. We now want to find conditions that need to hold for these heuristic approaches to succeed in spanning the SDF.

II.B. Heuristic factor construction: OLS factors and transformations thereof

Many heuristic methods construct factors by taking long positions in stocks with high values of a characteristic and short positions in stocks with low values of a characteristic, with the portfolio weight matrix and factors then taking the form

$$\boldsymbol{W}_t = \boldsymbol{X}_t \boldsymbol{S}_t, \qquad \boldsymbol{f}_{t+1} = \boldsymbol{W}_t' \boldsymbol{z}_{t+1}, \tag{14}$$

for some nonsingular matrix $\boldsymbol{S}_t$. For example, $\boldsymbol{S}_t = \boldsymbol{I}$ yields univariate factors with weights that are proportional to characteristics as, e.g., in Kozak, Nagel, and Santosh (2020). With characteristics defined as dummy variables for characteristics bins, portfolio sorts can also be represented in this way. Another example are cross-sectional regression slope factors. Fama and French (2020) use the insight of Fama (1976) that OLS cross-sectional regression slopes are themselves portfolio returns. This is the case $\boldsymbol{S}_t = (\boldsymbol{X}_t' \boldsymbol{X}_t)^{-1}$.

Fama and French (2020) conjecture that the OLS factors yield an "asset pricing model that can be used in time-series applications." In other words, they conjecture that for N assets with OLS factor betas $\boldsymbol{\beta}_t$, the pricing relation $\boldsymbol{\mu}_t = \boldsymbol{\beta}_t \boldsymbol{\mu}_{f,t}$ holds. However, such a pricing relationship does not generally hold for OLS factors. As we show now, this is true only if the covariance matrix takes a special form.

Proposition 2 *Suppose Assumption 1 holds and let $\boldsymbol{W}_t = \boldsymbol{X}_t \boldsymbol{S}_t$. Then, for any nonsingular $J \times J$ matrix $\boldsymbol{S}_t$, the maximum squared conditional Sharpe ratio of the factors $\boldsymbol{f}_{t+1} = \boldsymbol{W}_t' \boldsymbol{z}_{t+1}$ is equal to the maximum squared conditional Sharpe Ratio of the individual assets if and only if there exist conformable matrices $\boldsymbol{\Psi}_t$, $\boldsymbol{\Omega}_t$, and a matrix $\boldsymbol{U}_t$ for which*

$$\boldsymbol{U}_t' \boldsymbol{X}_t = \boldsymbol{0}, \tag{15}$$

such that

$$\boldsymbol{\Sigma}_t = \boldsymbol{X}_t \boldsymbol{\Psi}_t \boldsymbol{X}_t' + \boldsymbol{U}_t \boldsymbol{\Omega}_t \boldsymbol{U}_t'. \tag{16}$$

Proof. Lu and Schmidt (2012) Theorem 1 (B, F') implies that (16) is equivalent to the statement that there exists a nonsingular $\boldsymbol{B}_t$ such that $\boldsymbol{\Sigma}_t \boldsymbol{X}_t = \boldsymbol{X}_t \boldsymbol{B}_t$. Rewriting Assumption 1 as $\boldsymbol{\mu}_t = \boldsymbol{X}_t \boldsymbol{B}_t \boldsymbol{B}_t^{-1} \boldsymbol{\phi}$, we see that it is then equivalent to $\boldsymbol{\mu}_t = \boldsymbol{\Sigma}_t \boldsymbol{X}_t \boldsymbol{S}_t \boldsymbol{S}_t^{-1} \boldsymbol{B}_t^{-1} \boldsymbol{\phi} = \boldsymbol{\Sigma}_t \boldsymbol{W}_t \boldsymbol{b}_t$, where $\boldsymbol{b}_t = \boldsymbol{S}_t^{-1} \boldsymbol{B}_t^{-1} \boldsymbol{\phi}$. Thus, condition (8) in Lemma 1 is satisfied, which means that Lemma 1 applies. ■

Without the restriction (15), the decomposition in (16) would always exist. For instance, for any nonsingular symmetric $\boldsymbol{\Psi}_t$, we could obtain $\boldsymbol{U}_t \boldsymbol{\Omega}_t \boldsymbol{U}_t'$ from an eigendecomposition of $\boldsymbol{\Sigma}_t - \boldsymbol{X}_t \boldsymbol{\Psi}_t \boldsymbol{X}_t'$, where $\boldsymbol{U}_t$ then contains the eigenvectors associated with the $N - J$ nonzero eigenvalues in the diagonal matrix $\boldsymbol{\Omega}_t$.

How can researchers wishing to use OLS factors, or transformations thereof, ensure that the condition $\boldsymbol{U}_t' \boldsymbol{X}_t = \boldsymbol{0}$ in (15) holds, at least approximately? Including many characteristics in $\boldsymbol{X}_t$ should help. To see this, we can use the result in Lu and Schmidt (2012) that the conditions in (15) and (16) are equivalent to J eigenvectors of $\boldsymbol{\Sigma}_t$ being spanned by $\boldsymbol{X}_t$. The matrix $\boldsymbol{U}_t$ then contains linear combinations of the eigenvectors not spanned by $\boldsymbol{X}_t$.[5] With only a few characteristics included in $\boldsymbol{X}_t$, it is unlikely that the J columns of $\boldsymbol{X}_t$ exactly span J eigenvectors. Effectively, for each eigenvector, this is like asking whether a regression of the N elements of the eigenvector on the J variables in $\boldsymbol{X}_t$ has perfect fit. Clearly, the more characteristics we add, the better the fit. In this sense, it is more likely that $\boldsymbol{U}_t' \boldsymbol{X}_t = \boldsymbol{0}$ holds if $\boldsymbol{X}_t$ contains more characteristics.

Moreover, with a larger number of characteristics it is more likely that $\boldsymbol{X}_t$ spans very well the relatively small number of eigenvectors associated with large eigenvalues, i.e., the major sources of stock return covariance. In this case, even if $\boldsymbol{X}_t$ does not span J eigenvectors

5. If $\boldsymbol{Q}_t$ and $\boldsymbol{\Lambda}_t$ are the matrix of eigenvectors and diagonal matrix of eigenvalues of $\boldsymbol{\Sigma}_t$, respectively, and $\boldsymbol{Q}_t = (\boldsymbol{X}_t \boldsymbol{B}_t : \boldsymbol{U}_t)$ where the columns of $\boldsymbol{X}_t \boldsymbol{B}_t$, with nonsingular $\boldsymbol{B}_t$, are the J eigenvectors spanned by $\boldsymbol{X}_t$ and $\boldsymbol{U}_t$ are the eigenvectors not spanned by $\boldsymbol{X}_t$, then we have

$$\boldsymbol{\Sigma}_t = \boldsymbol{X}_t \boldsymbol{B}_t \boldsymbol{\Lambda}_{1,t} \boldsymbol{B}_t' \boldsymbol{X}_t' + \boldsymbol{U}_t \boldsymbol{\Lambda}_{2,t} \boldsymbol{U}_t' \tag{17}$$

which maps into (16) with $\boldsymbol{B}_t \boldsymbol{\Lambda}_{1,t} \boldsymbol{B}_t' = \boldsymbol{\Psi}_t$ and $\boldsymbol{\Lambda}_{2,t} = \boldsymbol{\Omega}_t$. Moreover, since eigenvectors are orthogonal, $\boldsymbol{B}_t' \boldsymbol{X}_t' \boldsymbol{U}_t = \boldsymbol{0}$ and hence $\boldsymbol{U}_t' \boldsymbol{X}_t = \boldsymbol{0}$.

perfectly, spanning the few important ones very well may render the violations of $\boldsymbol{U}_t'\boldsymbol{X}_t = \boldsymbol{0}$ quantitatively unimportant. OLS factors, or transformations thereof, may then span the SDF approximately. We investigate this further in our empirical analysis in Section V.

Importantly, for additional characteristics to be helpful in ensuring that $\boldsymbol{U}_t'\boldsymbol{X}_t = \boldsymbol{0}$ holds approximately, these additional characteristics do not necessarily need to contribute to variation in expected returns. If they help to span major sources of covariances, they will help OLS factors, or transformations thereof, to span the SDF, even without contribution to variation in expected returns.

The choice of transformation matrix $\boldsymbol{S}_t$ is again just one of convenience. Our first example shows the choice that yields OLS cross-sectional regression slope factors:

Example 3 *Suppose* $\boldsymbol{S}_t = (\boldsymbol{X}_t'\boldsymbol{X}_t)^{-1}$ *and that* (16) *holds. We then obtain an SDF with factors given by OLS cross-sectional regression slopes,* $\boldsymbol{f}_{t+1} = (\boldsymbol{X}_t'\boldsymbol{X}_t)^{-1}\boldsymbol{X}_t'\boldsymbol{z}_{t+1}$*. Prices of risk are* $\boldsymbol{b}_t = \boldsymbol{\Psi}_t^{-1}\boldsymbol{X}_t\boldsymbol{\phi}$ *and factor risk premia are constant,* $\boldsymbol{\mu}_{f,t} = \boldsymbol{\phi}$*. Factor betas are equal to characteristics,* $\boldsymbol{\beta}_t = \boldsymbol{X}_t$*.*

Based on this example, condition (16) can be interpreted through the lens of factor models. First note that the condition is equivalent to[6]

$$\boldsymbol{\Sigma}_t = \boldsymbol{X}_t\boldsymbol{\Psi}_t\boldsymbol{X}_t' + \boldsymbol{U}_t\boldsymbol{\Omega}_t\boldsymbol{U}_t' + \sigma^2\boldsymbol{I}, \qquad \boldsymbol{U}_t'\boldsymbol{X}_t = \boldsymbol{0}, \tag{18}$$

where we abuse notation a bit since $\boldsymbol{\Psi}_t$, $\boldsymbol{\Omega}_t$, and $\boldsymbol{U}_t$ are different here from (16). In the case of OLS factors this condition is equivalent to the factor model

$$\boldsymbol{z}_{t+1} = \boldsymbol{\alpha}_t + \boldsymbol{\beta}_{f,t}\left(\boldsymbol{f}_{t+1} - \mathbb{E}_t\,\boldsymbol{f}_{t+1}\right) + \boldsymbol{\beta}_{g,t}\left(\boldsymbol{g}_{t+1} - \mathbb{E}_t\,\boldsymbol{g}_{t+1}\right) + \boldsymbol{\varepsilon}_{t+1}, \tag{19}$$

where $\boldsymbol{f}_{t+1}$ are the OLS factors, with $\boldsymbol{\beta}_{f,t} = \boldsymbol{X}_t$ as in Example 3, $\boldsymbol{g}_{t+1}$ are latent factors, with $\boldsymbol{\beta}_{g,t} = \boldsymbol{U}_t$, and $\boldsymbol{\varepsilon}_{t+1}$ is a vector of idiosyncratic shocks. The condition (18) then requires

6. See Lu and Schmidt (2012) Theorem 1(F) and 1(F').

that latent risks not captured by the OLS factors are either idiosyncratic and homoskedastic, or, if they are systematic, individual assets' loadings on the latent factors are orthogonal to the loadings on the OLS factors. This condition is often assumed up front in factor models (e.g., in Kelly, Pruitt, and Su 2019).

A different choice of $\boldsymbol{S}_t$ produces univariate factors where weights in each characteristics portfolio depend only on one characteristic because the columns of $\boldsymbol{X}_t$ serve as weights:

Example 4 *Suppose* $\boldsymbol{S}_t = \boldsymbol{I}$ *and that* (16) *holds. We then obtain an SDF with factors* $\boldsymbol{f}_{t+1} = \boldsymbol{X}_t'\boldsymbol{z}_{t+1}$. *Factor risk prices,* $\boldsymbol{b}_t = (\boldsymbol{X}_t'\boldsymbol{X}_t)^{-1}\boldsymbol{\Psi}_t^{-1}\boldsymbol{\phi}$, *and factor means,* $\boldsymbol{\mu}_{f,t} = \boldsymbol{X}_t'\boldsymbol{X}_t\boldsymbol{\phi}$, *are time-varying. Factor betas are* $\boldsymbol{\beta}_t = \boldsymbol{X}_t(\boldsymbol{X}_t'\boldsymbol{X}_t)^{-1}$.

The special case in the latter example is particularly convenient for illustrating the meaning of $\boldsymbol{U}_t'\boldsymbol{X}_t = \boldsymbol{0}$ in (15). For the factor model to price assets perfectly, factor covariances must span $\boldsymbol{\mu}_t = \boldsymbol{X}_t\boldsymbol{\phi}$. This is always the case for the factors in Example 2 where factor weights are $\boldsymbol{\Sigma}_t^{-1}\boldsymbol{X}_t$ and hence individual assets' factor covariances are $\boldsymbol{\Sigma}_t\boldsymbol{\Sigma}_t^{-1}\boldsymbol{X}_t = \boldsymbol{X}_t$. In contrast, in the case of Example 4, individual assets' covariances with factor portfolios with factor weights $\boldsymbol{X}_t$ are

$$\boldsymbol{\Sigma}_t\boldsymbol{X}_t = \boldsymbol{X}_t\boldsymbol{\Psi}_t\boldsymbol{X}_t'\boldsymbol{X}_t + \boldsymbol{U}_t\boldsymbol{\Omega}_t\boldsymbol{U}_t'\boldsymbol{X}_t. \tag{20}$$

If $\boldsymbol{U}_t'\boldsymbol{X}_t = \boldsymbol{0}$, then the second component is zero and the expression hence simplifies to a term $\boldsymbol{X}_t$ multiplied by a nonsingular matrix. Factor covariances therefore span $\boldsymbol{\mu}_t$. But if $\boldsymbol{U}_t'\boldsymbol{X}_t \neq \boldsymbol{0}$ then the second component does not disappear. As a consequence, the factor covariances are contaminated by components that are not linear in $\boldsymbol{X}_t$, and hence are unpriced as they do not earn expected return. Thus, when $\boldsymbol{U}_t'\boldsymbol{X}_t \neq \boldsymbol{0}$, the factors with weights $\boldsymbol{X}_t$ incorporate unpriced risks, while factors that span the SDF capture only priced risks.

II.C. Hedged heuristic factors

If condition $\boldsymbol{U}_t'\boldsymbol{X}_t = \boldsymbol{0}$ in Proposition 2 does not hold, any factors with weights that are a nonsingular transformation of $\boldsymbol{X}_t$ load on unpriced risks, i.e., risk exposure that is not compensated with higher excess returns. This prevents the factors from reaching the mean-variance frontier.

Using the GLS factors, or transformations thereof, following Proposition 1 would avoid contamination of factors with unpriced risks, but their construction requires inversion of the large covariance matrix $\boldsymbol{\Sigma}_t$ (that would have to be estimated in practice). For this reason, it is useful to ask whether there exist an alternative factor specifications that use *some* information about covariances to find characteristics-based factors that span the SDF, but without requiring estimation and inversion of the whole covariance matrix $\boldsymbol{\Sigma}_t$. These factors will be *hedged* factors because they hedge unpriced exposures of the original factors.

We first show a result that will be helpful for checking whether a candidate hedged factor model with factor portfolio weight matrix $\boldsymbol{H}_t$ spans the SDF.

Lemma 2 *Suppose Assumption 1 holds and that $\boldsymbol{H}_t$ is some matrix such that $\boldsymbol{H}_t'\boldsymbol{X}_t$ has full column rank and $\boldsymbol{H}_t'\boldsymbol{\Sigma}_t\boldsymbol{H}_t$ is positive definite. Then the maximum squared conditional Sharpe ratio of the factors $\boldsymbol{f}_{t+1} = \boldsymbol{H}_t'\boldsymbol{z}_{t+1}$ is equal to the maximum conditional squared Sharpe Ratio of the individual assets if and only if there exist a nonsingular matrix $\boldsymbol{\Psi}_t$, and some matrices $\boldsymbol{\Omega}_t$ and $\boldsymbol{U}_t$ for which*

$$\boldsymbol{U}_t'\boldsymbol{H}_t = \boldsymbol{0}, \tag{21}$$

such that

$$\boldsymbol{\Sigma}_t = \boldsymbol{X}_t\boldsymbol{\Psi}_t\boldsymbol{X}_t' + \boldsymbol{U}_t\boldsymbol{\Omega}_t\boldsymbol{U}_t'. \tag{22}$$

Proof. Lu and Schmidt (2012) Theorem 3 (B, F') implies that (22) is equivalent to the statement that there exists some $\boldsymbol{B}_t$ such that $\boldsymbol{X}_t = \boldsymbol{\Sigma}_t\boldsymbol{H}_t\boldsymbol{B}_t$. Rewriting Assumption 1 as $\boldsymbol{\mu}_t = \boldsymbol{\Sigma}_t\boldsymbol{H}_t\boldsymbol{B}_t\boldsymbol{\phi} = \boldsymbol{\Sigma}_t\boldsymbol{W}_t\boldsymbol{b}_t$, where $\boldsymbol{b}_t = \boldsymbol{B}_t\boldsymbol{\phi}$. Thus, condition (8) is satisfied, which means that Lemma 1 applies. ■

There are two key points to note. First, the requirement that $\boldsymbol{H}_t'\boldsymbol{X}_t$ has full column rank ensures that no information about expected returns is lost when individual assets are aggregated with $\boldsymbol{H}_t$ as portfolio weight matrix. Second, the requirement that $\boldsymbol{U}_t'\boldsymbol{H}_t = \boldsymbol{0}$ ensures that the factors do not load on unpriced risk. When both conditions hold, a similar calculation as in (20) for the $\boldsymbol{S}_t = \boldsymbol{I}$ case, but now with $\boldsymbol{H}_t$ as factor portfolio weights yields

$$\boldsymbol{\Sigma}_t \boldsymbol{H}_t = \boldsymbol{X}_t \boldsymbol{\Psi}_t \boldsymbol{X}_t' \boldsymbol{H}_t + \boldsymbol{U}_t \boldsymbol{\Omega}_t \boldsymbol{U}_t' \boldsymbol{H}_t = \boldsymbol{X}_t \boldsymbol{\Psi}_t \boldsymbol{X}_t' \boldsymbol{H}_t, \tag{23}$$

which means that the individual assets' covariances with these factors are perfectly linear in $\boldsymbol{X}_t$ and so they span $\boldsymbol{\mu}_t$.

While Lemma 2 allows us to check whether candidate factors span the SDF, it does not show how to construct factors that satisfy these requirements. Some conditions on $\boldsymbol{U}_t \boldsymbol{\Omega}_t \boldsymbol{U}_t'$ will have to hold for the construction to be possible without using the information from the full $\boldsymbol{\Sigma}_t$ matrix. To see how additional structure on $\boldsymbol{U}_t \boldsymbol{\Omega}_t \boldsymbol{U}_t'$ can help, suppose that J columns of $\boldsymbol{U}_t$, collected in $\boldsymbol{V}_t$, are such that $\text{rank}(\boldsymbol{V}_t' \boldsymbol{X}_t) = J$, while the remaining columns, collected in $\boldsymbol{E}_t$, have $\boldsymbol{E}_t' \boldsymbol{X}_t = \boldsymbol{0}$ and $\boldsymbol{E}_t' \boldsymbol{V}_t = \boldsymbol{0}$. Moreover, suppose that $\boldsymbol{\Omega}_t$ is block-diagonal such that

$$\boldsymbol{U}_t \boldsymbol{\Omega}_t \boldsymbol{U}' = \boldsymbol{V}_t \boldsymbol{\Gamma}_t \boldsymbol{V}_t' + \boldsymbol{E}_t \boldsymbol{\Phi}_t \boldsymbol{E}_t'. \tag{24}$$

If we knew $\boldsymbol{V}_t$, we could then simply remove from characteristics-based factor weights $\boldsymbol{W}_t = \boldsymbol{X}_t \boldsymbol{S}_t$ the component that is correlated with $\boldsymbol{U}_t$ by subtracting the projection of the weights on $\boldsymbol{V}_t$,

$$\boldsymbol{H}_t = \boldsymbol{X}_t \boldsymbol{S}_t - \boldsymbol{V}_t (\boldsymbol{V}_t' \boldsymbol{V}_t)^{-1} \boldsymbol{V}_t' \boldsymbol{X}_t \boldsymbol{S}_t. \tag{25}$$

It is easy to verify that $\boldsymbol{H}_t' \boldsymbol{U}_t = \boldsymbol{0}$ and that $\boldsymbol{H}_t' \boldsymbol{X}_t$ has full column rank, i.e., the conditions in Lemma 2 hold.

We cannot directly implement this approach as $\boldsymbol{V}_t$ is not directly observable. But it can be backed out from moments of $\boldsymbol{z}_t$ and $\boldsymbol{X}_t$. As we show, the factor hedging method of Daniel, Mota, Rottke, and Santos (2020) (DMRS) is a feasible version of the approach above.

The goal of DMRS's procedure is to hedge the unpriced risk in heuristic factors. The first step is to construct hedging factors that go long in stocks with high loadings on the heuristic factors and short in stocks with low loadings, while holding constant the characteristics-exposure of the long and short legs of hedging factors, which ensures that they have zero expected return according to Assumption 1. DMRS do this by sorting stocks by loadings on heuristic factors within characteristics-sorted portfolios. Here, we work with more general characteristics-based factors with weights $\boldsymbol{W}_t = \boldsymbol{X}_t\boldsymbol{S}_t$ and we construct a hedging portfolio that has precisely zero expected return by regressing conditional covariances of individual stocks with factors, i.e., $\boldsymbol{\Sigma}_t\boldsymbol{W}_t$, on $\boldsymbol{X}_t$, and then using the residuals,

$$\boldsymbol{W}_{h,t} = \boldsymbol{R}_t\boldsymbol{\Sigma}_t\boldsymbol{X}_t\boldsymbol{S}_t \tag{26}$$

as portfolio weights for hedge portfolios.

The second step is to calculate stocks' covariances with the hedge portfolio returns so that we can modify stocks' weights in the factor portfolios to remove unpriced risks:

$$\hat{\boldsymbol{V}}_t = \boldsymbol{\Sigma}_t\boldsymbol{W}_{h,t} = \boldsymbol{V}_t\boldsymbol{\Gamma}_t\boldsymbol{V}_t'\boldsymbol{R}_t\boldsymbol{V}_t\boldsymbol{\Gamma}_t\boldsymbol{V}_t'\boldsymbol{X}_t\boldsymbol{S}_t. \tag{27}$$

The third step is to regress the factor portfolio weights $\boldsymbol{W}_t = \boldsymbol{X}_t\boldsymbol{S}_t$ on $\hat{\boldsymbol{V}}_t$ to obtain residual factor portfolio weights that have been purged of unpriced risk exposure. Now note that $\hat{\boldsymbol{V}}_t$ in (27) is equal to $\boldsymbol{V}_t$ post-multiplied by a nonsingular matrix. Hence regressing $\boldsymbol{W}_t$ on $\hat{\boldsymbol{V}}_t$ produces the same residuals as regressing $\boldsymbol{W}_t$ on $\boldsymbol{V}_t$. Therefore, the residuals[7]

$$\hat{\boldsymbol{H}}_t = \boldsymbol{X}_t\boldsymbol{S}_t - \hat{\boldsymbol{V}}_t(\hat{\boldsymbol{V}}_t'\hat{\boldsymbol{V}}_t)^{-1}\hat{\boldsymbol{V}}_t'\boldsymbol{X}_t\boldsymbol{S}_t \tag{30}$$

are the same as the residuals in (25) and hence $\hat{\boldsymbol{H}}_t = \boldsymbol{H}_t$. In other words, the three steps

7. DMRS use a slightly different approach, but under the assumptions of Proposition 3 below, it yields the same hedged factors. They purge the heuristic factors from unpriced risks that do not earn expected return by regressing the J heuristic factors on the J hedge portfolio returns and using the J time series of residuals as the hedged factors. The $J \times J$ matrix of regression coefficients in these regressions is

$$\boldsymbol{K}_t = \boldsymbol{S}_t'\boldsymbol{X}_t'\boldsymbol{W}_{h,t}(\boldsymbol{W}_{h,t}'\boldsymbol{\Sigma}_t\boldsymbol{W}_{h,t})^{-1}\boldsymbol{S}_t^{-1}, \tag{28}$$

and so the hedged factors have weights

$$\hat{\boldsymbol{H}}_t = \boldsymbol{X}_t\boldsymbol{S}_t - \boldsymbol{W}_{h,t}\boldsymbol{K}_t'. \tag{29}$$

Substituting $\boldsymbol{W}_{h,t} = \boldsymbol{V}_t\boldsymbol{A}_t$, for some nonsingular matrix $\boldsymbol{A}_t$, into this expression and $\boldsymbol{K}_t$, it can be seen that this last expression is equivalent to (25).

above provide a way to construct the hedged portfolio weights in (25) from observable moments.

The following proposition states the result more formally.

Proposition 3 *If the matrices $\boldsymbol{U}_t$ and $\boldsymbol{\Omega}_t$ in* (22) *are such that there exists a decomposition*

$$\boldsymbol{U}_t\boldsymbol{\Omega}_t\boldsymbol{U}_t' = \boldsymbol{V}_t\boldsymbol{\Gamma}_t\boldsymbol{V}_t' + \boldsymbol{E}_t\boldsymbol{\Phi}_t\boldsymbol{E}_t', \tag{31}$$

where $\boldsymbol{V}_t$ is an $N \times J$ matrix of full column rank, $\boldsymbol{V}_t'\boldsymbol{X}_t$ is full rank, $\boldsymbol{R}_t\boldsymbol{V}_t$ has full column rank, $\boldsymbol{E}_t'\boldsymbol{X}_t = \boldsymbol{0}$, $\boldsymbol{E}_t'\boldsymbol{V}_t = \boldsymbol{0}$ and $\boldsymbol{\Gamma}_t$ is nonsingular, then the maximum squared conditional Sharpe ratio of the hedged factors $\boldsymbol{f}_{t+1} = \hat{\boldsymbol{H}}_t'\boldsymbol{z}_{t+1}$ with $\hat{\boldsymbol{H}}_t$ as defined in (30) *is equal to the maximum squared conditional Sharpe Ratio of the individual assets.*

Proof. Write $\hat{\boldsymbol{V}}_t = \boldsymbol{V}_t\boldsymbol{A}_t$ where $\boldsymbol{A}_t = \boldsymbol{\Gamma}_t\boldsymbol{V}_t'\boldsymbol{R}_t\boldsymbol{V}_t\boldsymbol{\Gamma}_t\boldsymbol{V}_t'\boldsymbol{X}_t\boldsymbol{S}_t$. By assumption, $\boldsymbol{R}_t\boldsymbol{V}_t$ has full column rank J, hence $\boldsymbol{V}_t'\boldsymbol{R}_t\boldsymbol{V}_t = \boldsymbol{V}_t'\boldsymbol{R}_t\boldsymbol{R}_t\boldsymbol{V}_t$ has full rank. Since pre- and post-multiplying this expression by full rank matrices $\boldsymbol{\Gamma}_t$ and $\boldsymbol{V}_t'\boldsymbol{X}_t\boldsymbol{S}_t$ does not change rank, it follows that $\boldsymbol{A}_t$ is full rank and hence nonsingular. Then, substituting $\hat{\boldsymbol{V}}_t = \boldsymbol{V}_t\boldsymbol{A}_t$, with $\boldsymbol{A}_t$ nonsingular into (30) yields the expression for $\boldsymbol{H}_t$ in (25), i.e., $\hat{\boldsymbol{H}}_t = \boldsymbol{H}_t$. Then $\boldsymbol{U}_t'\hat{\boldsymbol{H}}_t = \boldsymbol{0}$ immediately follows. Therefore, by Lemma 2, the result follows. ■

The rank requirements for several matrices in Proposition 3 have an economic interpretation. That $\boldsymbol{V}_t'\boldsymbol{X}_t$ has full rank and $\boldsymbol{R}_t\boldsymbol{V}_t$ has full column rank ensures that the hedging portfolio weight vectors constructed via (27) and (30) are linearly independent. One could relax these rank requirements by building in a dimension-reduction step that removes linear dependencies in the construction of $\boldsymbol{W}_{h,t}$. However, for our purposes here, the benefits from greater generality of this approach would not be worth the costs of additional expositional complexity.

What do we gain from the hedging procedure? Comparing the conditions in Proposition 3 with (15) and (16) in Proposition 2, we can see that the conditions on the covariance matrix that are required to hold for the hedged factors to span the SDF are weaker than those required for the OLS factors (or nonsingular transformations thereof) to span the SDF. While Proposition 2 requires the columns of $\boldsymbol{X}_t$ to be orthogonal to the columns of $\boldsymbol{U}_t$, the conditions in Proposition 3 allow violations of this orthogonality condition as long as there are at most J linearly independent sources of such non-orthogonality as collected in the J columns of the matrix $\boldsymbol{V}_t$.

II.D. Iterated hedging

When $\boldsymbol{V}_t$ has more than J columns, then the (infeasible) hedged factor construction based on the unobservable $\boldsymbol{V}_t$ as in (25) still works as $\boldsymbol{H}_t'\boldsymbol{U}_t = \boldsymbol{0}$ still holds and $\boldsymbol{H}_t'\boldsymbol{X}_t$ still has full column rank, i.e., the conditions in Lemma 2 still hold. However, in this case the feasible hedged factor weights $\hat{\boldsymbol{H}}_t$ we construct in (30) are no longer equal to $\boldsymbol{H}_t$. The reason is that if we again construct $\hat{\boldsymbol{V}}_t$ as in (27), the J columns of $\hat{\boldsymbol{V}}_t$ now contain J linear combinations of the $2J$ columns in $\boldsymbol{V}_t$. Projection on $\hat{\boldsymbol{V}}_t$ therefore no longer produces the same residuals as a projection on $\boldsymbol{V}_t$.

However, by iterating on the hedging procedure, we can solve this problem. Repeating the hedging procedure by regressing individual stocks' conditional covariances with *hedged* factors, i.e., $\boldsymbol{\Sigma}_t\hat{\boldsymbol{H}}_t$, on $\boldsymbol{X}_t$ and collecting the residuals $\boldsymbol{R}_t\boldsymbol{\Sigma}_t\hat{\boldsymbol{H}}_t$ analogous to (26), but here for hedged factors. Using these residuals as portfolio weights, and calculating the covariances of individual stocks with these portfolio returns, we get, in analogy to (27),

$$\hat{\boldsymbol{V}}_{2,t} = \boldsymbol{V}_t\boldsymbol{\Gamma}_t\boldsymbol{V}_t'\boldsymbol{R}_t\boldsymbol{V}_t\boldsymbol{\Gamma}_t\boldsymbol{V}_t'\hat{\boldsymbol{H}}_t, \tag{32}$$

where the only difference to (27) is that $\boldsymbol{X}_t\boldsymbol{S}_t$ was replaced by $\hat{\boldsymbol{H}}_t$. Note that $\hat{\boldsymbol{V}}_{2,t}$ is comprised of J linear combinations of the $2J$ columns of $\boldsymbol{V}_t$.

Under conditions that we state more formally shortly, $\hat{\boldsymbol{V}}_t$ and $\hat{\boldsymbol{V}}_{2,t}$ jointly span the same column space as $\boldsymbol{V}_t$. Therefore, the residuals from the regression of $\boldsymbol{X}_t\boldsymbol{S}_t$ on $\boldsymbol{V}_t$ in (25) are the same as those from a regression of $\boldsymbol{X}_t\boldsymbol{S}_t$ on $\hat{\boldsymbol{V}}_t$ and $\hat{\boldsymbol{V}}_{2,t}$ jointly. And the latter regression can in turn be implemented in two steps, which results in an iterated hedging procedure. By the Frisch-Waugh-Lovell theorem, the residuals of a regression of $\boldsymbol{X}_t\boldsymbol{S}_t$ on $\hat{\boldsymbol{V}}_t$ and $\hat{\boldsymbol{V}}_{2,t}$ jointly are the same as the residuals of a regression of the first step residuals $\hat{\boldsymbol{H}}_t$

from regressing $\boldsymbol{X}_t\boldsymbol{S}_t$ on $\hat{\boldsymbol{V}}_t$ in (30) on the residuals from regressing $\hat{\boldsymbol{V}}_{2,t}$ on $\hat{\boldsymbol{V}}_t$. Therefore, we can construct the hedged portfolio weights as

$$\begin{aligned}\hat{\boldsymbol{H}}_{2,t} &= \boldsymbol{M}_t\boldsymbol{X}_t\boldsymbol{S}_t - \boldsymbol{M}_t\hat{\boldsymbol{V}}_{2,t}(\hat{\boldsymbol{V}}_{2,t}'\boldsymbol{M}_t\hat{\boldsymbol{V}}_{2,t})^{-1}\hat{\boldsymbol{V}}_{2,t}'\boldsymbol{M}_t\boldsymbol{X}_t\boldsymbol{S}_t \\ &= \hat{\boldsymbol{H}}_t - \boldsymbol{M}_t\hat{\boldsymbol{V}}_{2,t}(\hat{\boldsymbol{V}}_{2,t}'\boldsymbol{M}_t\hat{\boldsymbol{V}}_{2,t})^{-1}\hat{\boldsymbol{V}}_{2,t}'\boldsymbol{M}_t\hat{\boldsymbol{H}}_t, \end{aligned} \tag{33}$$

where $\boldsymbol{M}_t = \boldsymbol{I} - \hat{\boldsymbol{V}}_t(\hat{\boldsymbol{V}}_t'\hat{\boldsymbol{V}}_t)^{-1}\hat{\boldsymbol{V}}_t'$ is the residual maker matrix from regression on $\hat{\boldsymbol{V}}_t$, and we obtain $\hat{\boldsymbol{H}}_{2,t} = \boldsymbol{H}_t$.

The following proposition states this result formally. It looks similar to Proposition 3, but note that $\boldsymbol{V}_t$ now has $2J$ columns.

Proposition 4 *If the matrices $\boldsymbol{U}_t$ and $\boldsymbol{\Omega}_t$ in* (22) *are such that there exists a decomposition*

$$\boldsymbol{U}_t\boldsymbol{\Omega}_t\boldsymbol{U}_t' = \boldsymbol{V}_t\boldsymbol{\Gamma}_t\boldsymbol{V}_t' + \boldsymbol{E}_t\boldsymbol{\Phi}_t\boldsymbol{E}_t', \tag{34}$$

where $\boldsymbol{V}_t$ is an $N \times 2J$ matrix of full column rank, $\boldsymbol{R}_t\boldsymbol{V}_t$ has full column rank, $(\boldsymbol{V}_t'\boldsymbol{X}_t\boldsymbol{S}_t : \boldsymbol{V}_t'\hat{\boldsymbol{H}}_t)$ has full rank, with $\hat{\boldsymbol{H}}_t$ defined as in (30)*, $\boldsymbol{E}_t'\boldsymbol{X}_t = \boldsymbol{0}$, $\boldsymbol{E}_t'\boldsymbol{V}_t = \boldsymbol{0}$ and $\boldsymbol{\Gamma}_t$ is nonsingular, then the maximum squared conditional Sharpe ratio of the hedged factors $\boldsymbol{f}_{t+1} = \hat{\boldsymbol{H}}_{2,t}'\boldsymbol{z}_{t+1}$ with $\hat{\boldsymbol{H}}_{2,t}$ as defined in* (33) *is equal to the maximum squared Sharpe Ratio of the individual assets.*

Proof. We first show $\hat{\boldsymbol{V}}_t$ and $\hat{\boldsymbol{V}}_{2,t}$ jointly span the same column space as $\boldsymbol{V}_t$. Note that we can write $(\hat{\boldsymbol{V}}_t : \hat{\boldsymbol{V}}_{2,t}) = \boldsymbol{V}_t\boldsymbol{G}_t\boldsymbol{A}_t$ with $\boldsymbol{A}_t = (\boldsymbol{V}_t'\boldsymbol{X}_t\boldsymbol{S}_t : \boldsymbol{V}_t'\hat{\boldsymbol{H}}_t)$ where $\boldsymbol{G}_t = \boldsymbol{\Gamma}_t\boldsymbol{V}_t'\boldsymbol{R}_t\boldsymbol{V}_t\boldsymbol{\Gamma}_t$ is a full-rank $2J \times 2J$ square matrix ($\boldsymbol{R}_t\boldsymbol{V}_t$ has full column rank, so $\boldsymbol{R}_t\boldsymbol{V}_t\boldsymbol{\Gamma}_t$ has rank $2J$. Premultiplying $\boldsymbol{R}_t\boldsymbol{V}_t\boldsymbol{\Gamma}_t$ with its own transpose then results in a matrix that is also of rank $2J$). Since $\boldsymbol{A}_t$ and $\boldsymbol{G}_t$ are full rank and hence invertible, we have $\boldsymbol{V}_t = (\hat{\boldsymbol{V}}_t : \hat{\boldsymbol{V}}_{2,t})\boldsymbol{A}_t^{-1}\boldsymbol{G}_t^{-1}$, i.e., $\hat{\boldsymbol{V}}_t$ and $\hat{\boldsymbol{V}}_{2,t}$ jointly span the same column space as $\boldsymbol{V}_t$. Substituting this relation into (25), we obtain the residuals of a regression of $\boldsymbol{X}_t\boldsymbol{S}_t$ on $(\hat{\boldsymbol{V}}_t : \hat{\boldsymbol{V}}_{2,t})$. By the Frisch-Waugh-Lovell theorem, these residuals are in turn identical to those in the regression of $\boldsymbol{M}_t\boldsymbol{X}_t\boldsymbol{S}_t$ on $\boldsymbol{M}_t\hat{\boldsymbol{V}}_{2,t}$ in (33). Hence $\hat{\boldsymbol{H}}_{2,t} = \boldsymbol{H}_t$ and so $\hat{\boldsymbol{H}}_{2,t}'\boldsymbol{U}_t = \boldsymbol{0}$. Therefore, by Lemma 2, the result follows. ■

In analogy to the case with a single round of hedging that we discussed following Proposition 3, the rank requirements for several matrices in Proposition 4 have an economic interpretation. The requirements that $(\boldsymbol{V}_t'\boldsymbol{X}_t\boldsymbol{S}_t : \boldsymbol{V}_t'\hat{\boldsymbol{H}}_t)$ has full rank and $\boldsymbol{R}_t\boldsymbol{V}_t$ has full column rank are both needed to ensure that iterated hedging factor portfolio weight vectors $\hat{\boldsymbol{H}}_{2,t}$ are linearly independent. One could again relax these rank requirements by building dimension-reduction steps that removes linear dependencies in the iterated hedging procedure.

What do we gain from iterated hedging? Comparing the conditions in Proposition 4 with those in Proposition 3, we can see that those in Proposition 4 are weaker. While the conditions in Proposition 3 allow for J linearly independent sources of such non-orthogonality of $\boldsymbol{X}_t$ and the columns of $\boldsymbol{U}_t$, the conditions in Proposition 4 allow for $2J$ linearly independent sources of such non-orthogonality. In other words, iterated hedging can remove more sources of unpriced risk contamination in characteristics-based factors than a single round of hedging can.

There is no reason to necessarily stop after a second round of hedging. We do not show formal results on this, but from the logic of the hedging iteration above, it should be clear that further rounds of hedging would remove additional sources of unpriced risk contamination. When working with population moments, this should further raise the maximum squared conditional Sharpe Ratio of the hedged factors and hence get them closer to spanning the SDF. Whether this is also true in a finite sample with estimated moments is not clear. At some point, further hedging may be counterproductive and bring in estimation error contamination rather than removing unpriced risk contamination. After all, doing many iterations of the hedging procedure should be no different than constructing GLS factors by inverting an estimate of the conditional covariance matrix (which may not work well unless N is small relative to T). We investigate this further in Section V.

II.E. Summary

When conditional expected returns are linear in firm characteristics, aggregation of individual stocks into characteristics-based factor portfolios without incorporating information from the conditional covariance matrix of individual stock returns leads to a deterioration of the investment opportunity set unless the conditional covariance matrix satisfies certain conditions. These conditions are more likely to hold in large-scale factor models that use many characteristics. Methods for hedging unpriced risks in factors allow a partial relaxation of these conditions, especially if hedging procedures are applied iteratively.

III. Dimensionality reduction

So far we have discussed factor models where the pricing information in J characteristics is captured by J factors in the SDF. As we show now, under certain conditions on the conditional covariance matrix of individual stock returns, one can summarize the pricing information in J characteristics-based factors in a smaller number of $K < J$ factors. Of course, there is always a single factor that prices the individual assets (which the linear combination of J factors shown in Proposition 1), but without further assumptions, the construction of this single factor requires inversion of a large conditional covariance matrix. The point of the methods we discuss in this section is to achieve dimension reduction without having to invert or eigen-decompose this large covariance matrix.

We first present general conditions on the conditional covariance matrix that need to hold such that dimension reduction is possible without loss of pricing information. Then we show that, under these conditions, various approaches that have appeared in the literature are actually equivalent or closely related.

Corollary 2 *Suppose expected returns are given by*

$$\boldsymbol{\mu}_t = \boldsymbol{X}_t \boldsymbol{Q}_t \boldsymbol{\phi}, \tag{35}$$

where $\boldsymbol{Q}_t$ is a $J \times K$ matrix with $K \leq J$, and let $\boldsymbol{W}_t = \boldsymbol{X}_t \boldsymbol{Q}_t \boldsymbol{S}_t$. Then the maximum squared conditional Sharpe ratio of the factors $\boldsymbol{f}_{t+1} = \boldsymbol{W}_t' \boldsymbol{z}_{t+1}$, for any nonsingular $K \times K$ matrix $\boldsymbol{S}_t$, is equal to the maximum squared Sharpe Ratio of the individual assets if and only if there *exist conformable matrices* $\boldsymbol{\Lambda}_t$, $\boldsymbol{\Omega}_t$, *and a matrix* $\boldsymbol{U}_t$ *for which*

$$\boldsymbol{U}_t'\boldsymbol{X}_t\boldsymbol{Q}_t = \boldsymbol{0}, \tag{36}$$

such that

$$\boldsymbol{\Sigma}_t = \boldsymbol{X}_t\boldsymbol{Q}_t\boldsymbol{\Lambda}_t\boldsymbol{Q}_t'\boldsymbol{X}_t' + \boldsymbol{U}_t\boldsymbol{\Omega}_t\boldsymbol{U}_t'. \tag{37}$$

Proof. Directly follows from Proposition 2 by using $\boldsymbol{X}_t\boldsymbol{Q}_t$ in place of $\boldsymbol{X}_t$. ■

We have achieved dimension reduction because there are now K factors in $\boldsymbol{f}$, not J. This is made possible by the fact that the factor component of the covariance matrix related to $\boldsymbol{X}_t$ is now a lower-dimensional $\boldsymbol{X}_t\boldsymbol{Q}_t$, which is $N \times K$, with $K \leq J$, rather than the larger $N \times J$ matrix $\boldsymbol{X}_t$ that we had in Proposition 2. And $\boldsymbol{\Lambda}_t$ is a $K \times K$ matrix rather than the $J \times J$ matrix $\boldsymbol{\Psi}_t$ in Proposition 2.

How can we find $\boldsymbol{Q}_t$ to construct the factors $\boldsymbol{f}$? As we show now, if we make a somewhat stronger assumption than (36), namely that $\boldsymbol{U}_t'\boldsymbol{X}_t = \boldsymbol{0}$ we can obtain $\boldsymbol{Q}_t$ through principal component analysis (PCA). Under this assumption, OLS factors, for instance, and transformations thereof span the SDF. PCA applied to OLS factors can then extract $\boldsymbol{Q}_t$. More precisely, to extract $\boldsymbol{Q}_t$ as principal components, we need to add additional identification assumptions on $\boldsymbol{Q}_t$ and $\boldsymbol{\Lambda}_t$. These assumptions pin down a specific rotation of $\boldsymbol{Q}_t$, but they do not affect the pricing implications of the factor model. With different choices of identifying assumptions, we then obtain conditional versions of two recently proposed methods of dimension-reduced factor construction.

Example 5 *(IPCA) Suppose* $\boldsymbol{U}_t'\boldsymbol{X}_t = \boldsymbol{0}$, $\boldsymbol{Q}_t'\boldsymbol{Q}_t = \boldsymbol{I}$ *and* $\boldsymbol{\Lambda}_t$ *is diagonal with descending diagonal entries.*[8] *We can then obtain* $\boldsymbol{Q}_t$ *and* $\boldsymbol{\Lambda}_t$ *from an eigendecomposition of the conditional covariance matrix of OLS factor returns, because it factors as*

$$\left(\boldsymbol{X}_t'\boldsymbol{X}_t\right)^{-1}\boldsymbol{X}_t'\boldsymbol{\Sigma}_t\boldsymbol{X}_t\left(\boldsymbol{X}_t'\boldsymbol{X}_t\right)^{-1} = \boldsymbol{Q}_t\boldsymbol{\Lambda}_t\boldsymbol{Q}_t', \tag{38}$$

8. The last two assumptions correspond to identification assumption in Kelly, Pruitt, and Su (2019): $\boldsymbol{\Gamma}_\beta'\boldsymbol{\Gamma}_\beta = \boldsymbol{I}_K$ and $\text{cov}(\boldsymbol{f}_t)$ has only descending diagonal entries (their notation).

where $\boldsymbol{Q}_t$, given the assumptions above, becomes a matrix of eigenvectors of this covariance matrix associated with the K non-zero eigenvalues. Suppose further that $\boldsymbol{S}_t = (\boldsymbol{Q}'\boldsymbol{X}_t'\boldsymbol{X}_t\boldsymbol{Q})^{-1}$. Then we obtain a conditional version of the IPCA factors of Kelly, Pruitt, and Su (2019):

$$\boldsymbol{f}_{IPCA,t+1} = (\boldsymbol{Q}_t'\boldsymbol{X}_t'\boldsymbol{X}_t\boldsymbol{Q}_t)^{-1}\boldsymbol{Q}_t'\boldsymbol{X}_t'\boldsymbol{z}_{t+1}. \tag{39}$$

The expression for $\boldsymbol{f}_{IPCA,t+1}$ in (39) above is a conditional version of the first of two first-order conditions in Kelly, Pruitt, and Su (2019) that define the instrumented principal components analysis (IPCA) estimator. We can also show that a conditional version of their second first-order condition (their eq. 7) holds in terms of population moments. If it holds, then the right-hand side their second first-order condition should equal $\text{vec}(\boldsymbol{Q}_t)$ when evaluated with the factors in (39) and under the conditions of Corollary 2. Evaluating their second first-order condition, this is indeed what we obtain:

$$\begin{aligned}
&\left(\boldsymbol{X}_t'\boldsymbol{X}_t \otimes \mathbb{E}_t[\boldsymbol{f}_{t+1}\boldsymbol{f}_{t+1}']\right)^{-1} \mathbb{E}_t\left[\left(\boldsymbol{X}_t' \otimes \boldsymbol{f}_{t+1}\right)\boldsymbol{z}_{t+1}\right] \\
&= \left(\boldsymbol{X}_t'\boldsymbol{X}_t \otimes \mathbb{E}_t[\boldsymbol{f}_{t+1}\boldsymbol{f}_{t+1}']\right)^{-1} \text{vec}(\boldsymbol{E}_t[\boldsymbol{f}_{t+1}\boldsymbol{z}_{t+1}']\boldsymbol{X}_t) \\
&= \text{vec}\left(\mathbb{E}_t[\boldsymbol{f}_{t+1}\boldsymbol{f}_{t+1}']^{-1}\boldsymbol{E}_t[\boldsymbol{f}_{t+1}\boldsymbol{z}_{t+1}']\boldsymbol{X}_t\left(\boldsymbol{X}_t'\boldsymbol{X}_t\right)^{-1}\right) \\
&= \text{vec}(\boldsymbol{Q}_t),
\end{aligned} \tag{40}$$

where for the last step we evaluated the conditional expectations using (39), (35), (37), and (36). Hence, factors constructed as in (39) with $\boldsymbol{Q}_t$ obtained as eigenvectors of the OLS factor return covariance matrix in (38) solve both first-order conditions, i.e., they are indeed the IPCA factors.[9]

Kelly, Pruitt, and Su (2019) show that in the case of orthonormalized characteristics, IPCA is equivalent to PCA on returns managed portfolios with weights $\boldsymbol{X}_t'$. Our result here

9. The assumption of time-constant $\boldsymbol{Q}_t$ and $\boldsymbol{\Lambda}_t$ can justify working with a constant $\boldsymbol{Q}$ extracted from an average conditional, or approximately unconditional, covariance matrix. Working through the first-order condition in (40) expressed in terms of unconditional expectations (the population analog to the sample

shows that IPCA is more generally equivalent to PCA on managed portfolios, even in the case where characteristics are not orthonormalized, if the managed portfolios are constructed as OLS factors. In particular, applying PCA to OLS portfolios recovers $\boldsymbol{Q}_t$. By applying this matrix to univariate portfolios $\boldsymbol{X}_t'\boldsymbol{z}_{t+1}$ and further transforming them by an OLS factor (as in (39)), yields our version of the IPCA estimator.

The OLS factor population covariance matrix that we apply PCA to in (38) is singular if $K < J$ as it is a $J \times J$ matrix with only K non-zero eigenvalues. The matrices $\boldsymbol{\Lambda}_t$ and $\boldsymbol{Q}_t$ in our notation contain only the non-zero eigenvalues and the eigenvectors associated with the non-zero eigenvalues. With an estimated covariance matrix in a finite sample, the truly zero eigenvalues would not be exactly zero but likely very small.

Example 6 *(PPCA) Suppose* $\boldsymbol{U}_t'\boldsymbol{X}_t = \boldsymbol{0}$, $\boldsymbol{Q}_t'\boldsymbol{X}_t'\boldsymbol{X}_t\boldsymbol{Q}_t = \boldsymbol{I}$ *and* $\boldsymbol{\Lambda}_t$ *is diagonal with descending diagonal entries.*[10] *We can then obtain* $\boldsymbol{Q}_t$ *and* $\boldsymbol{\Lambda}_t$ *from an eigendecomposition of the conditional covariance matrix of univariate factor returns constructed using orthonormalized characteristics, because it factors as*

$$\left(\boldsymbol{X}_t'\boldsymbol{X}_t\right)^{-\frac{1}{2}} \boldsymbol{X}_t'\boldsymbol{\Sigma}_t\boldsymbol{X}_t \left(\boldsymbol{X}_t'\boldsymbol{X}_t\right)^{-\frac{1}{2}} = \left(\boldsymbol{X}_t'\boldsymbol{X}_t\right)^{\frac{1}{2}} \boldsymbol{Q}_t\boldsymbol{\Lambda}_t\boldsymbol{Q}_t' \left(\boldsymbol{X}_t'\boldsymbol{X}_t\right)^{\frac{1}{2}}, \tag{41}$$

where $\boldsymbol{G}_t = (\boldsymbol{X}_t'\boldsymbol{X}_t)^{\frac{1}{2}}\boldsymbol{Q}_t$ *is orthonormal by assumption and thus can be recovered as a matrix of eigenvectors of this covariance matrix associated with the* K *non-zero eigenvalues. We get* $\boldsymbol{Q}_t = (\boldsymbol{X}_t'\boldsymbol{X}_t)^{-\frac{1}{2}}\boldsymbol{G}_t$. *Suppose further that* $\boldsymbol{S}_t = \boldsymbol{I}$. *Then we obtain a conditional version of*

averages in KPS), we the obtain vec($\boldsymbol{Q}$):

$$\begin{aligned}
&\left[\mathbb{E}\left(\boldsymbol{X}_t'\boldsymbol{X}_t \otimes \mathbb{E}_t\left[\boldsymbol{f}_{KPS,t+1}\boldsymbol{f}_{KPS,t+1}'\right]\right)\right]^{-1} \mathbb{E}\left[\left(\boldsymbol{X}_t' \otimes \boldsymbol{f}_{KPS,t+1}\right)\boldsymbol{z}_{t+1}\right] \\
&= \left[\mathbb{E}\left(\boldsymbol{X}_t'\boldsymbol{X}_t\right) \otimes \boldsymbol{\Lambda}\right]^{-1} \text{vec}\left(\mathbb{E}\left[\mathbb{E}_t[\boldsymbol{f}_{KPS,t+1}\boldsymbol{z}_{t+1}']\boldsymbol{X}_t\right]\right) \\
&= \left[\mathbb{E}\left(\boldsymbol{X}_t'\boldsymbol{X}_t\right) \otimes \boldsymbol{\Lambda}\right]^{-1} \text{vec}\left(\boldsymbol{\Lambda}\boldsymbol{Q}\,\mathbb{E}\left(\boldsymbol{X}_t'\boldsymbol{X}_t\right)\right) = \text{vec}(\boldsymbol{Q}).
\end{aligned}$$

10. The last two assumptions correspond to identification assumptions stated in assumption 3 of Kim, Korajczyk, and Neuhierl (2021). Our assumption that $\boldsymbol{U}'\boldsymbol{X}_t = \boldsymbol{0}$ is the population version of their assumption 2 (ii), which states that factor model residuals and $\boldsymbol{X}_t$ are, asymptotically, cross-sectionally orthogonal.

the PPCA factors of Kim, Korajczyk, and Neuhierl (2021):

$$\boldsymbol{f}_{PPCA,t+1} = \boldsymbol{G}_t' \left(\boldsymbol{X}_t'\boldsymbol{X}_t\right)^{-\frac{1}{2}} \boldsymbol{X}_t'\boldsymbol{z}_{t+1} \tag{42}$$

$$= \left(\boldsymbol{Q}_t'\boldsymbol{X}_t'\boldsymbol{X}_t\boldsymbol{Q}_t\right)^{-1} \boldsymbol{Q}_t'\boldsymbol{X}_t'\boldsymbol{z}_{t+1} = \boldsymbol{Q}_t'\boldsymbol{X}_t'\boldsymbol{z}_{t+1}. \tag{43}$$

The expression for $\boldsymbol{f}_{PPCA,t+1}$ in (43) above is a conditional version of the factors in Kim, Korajczyk, and Neuhierl (2021) obtained from a cross-sectional regression of stock returns on their factor loadings $\boldsymbol{G}_\beta(\boldsymbol{X}_t)$ which we parameterize as linear here, $\boldsymbol{G}_\beta(\boldsymbol{X}_t) = \boldsymbol{X}_t\boldsymbol{Q}_t$. To see this, note that Kim, Korajczyk, and Neuhierl (2021) identify $\boldsymbol{G}_\beta(\boldsymbol{X}_t)$ via a PCA on projected returns, $\boldsymbol{X}_t \left(\boldsymbol{X}_t'\boldsymbol{X}_t\right)^{-1} \boldsymbol{X}_t'\boldsymbol{z}_{t+1}$. Under our assumption in (37), the covariance matrix of these returns is equal to $\boldsymbol{X}_t\boldsymbol{Q}_t\boldsymbol{\Lambda}_t\boldsymbol{Q}_t'\boldsymbol{X}_t'$. Because $\boldsymbol{X}_t\boldsymbol{Q}_t$ is orthonormal, Kim et al.'s PCA solution, therefore, recovers $\boldsymbol{G}_\beta(\boldsymbol{X}_t) = \boldsymbol{X}_t\boldsymbol{Q}_t$ and their factors match ours in (43). The expression in (42) shows that we can alternatively identify these factors via a simple PCA on univariate portfolio returns (rather than projected individual stock returns) constructed using orthonormalized characteristics, to obtain $\boldsymbol{G}_t$.

Overall, the results in this section show that there is a great deal of similarity in seemingly different recently proposed methods for dimension reduction. Our earlier results on the conditions required for characteristics-based factors to span the SDF provide a basis to get to these dimension-reduction in a straightforward way by applying PCA to a certain set of characteristics-based portfolios.

IV. Extensions

Before turning to an empirical analysis, we first discuss a number of conceptual issues that come up if we want to relate our results from the previous sections to empirical data.

IV.A. *Alternative assumptions about expected returns*

As we discussed, our Assumption 1 that conditional expected returns are linear in characteristics is, in principle, completely general as for any given set of basis characteristics, one could define $\boldsymbol{X}_t$ as including nonlinear functions and interactions of these basis characteristics. That said, once a researcher has settled on a particular set of characteristics to include in $\boldsymbol{X}_t$, the linearity assumption has economic content. For this reason, one may want to entertain alternative assumptions that link a specific characteristics matrix $\boldsymbol{X}_t$ to $\boldsymbol{\mu}_t$.

For example, within a framework in which characteristics predict returns because of mispricing, our baseline Assumption 1 can be reasonable if the characteristics in $\boldsymbol{X}_t$ are directly related to the magnitude of mispricing without involving cross-asset information. As an example, consider scaled price ratios like the book-to-market ratio. If the numerator (book value) controls for differences across stocks in their fundamental scale and the remaining price variation that comes in through the denominator (market value), each stock's book-to-market ratio may be a good measure of this stock's mispricing.

However, an alternative view may be that characteristics in $\boldsymbol{X}_t$ capture not the magnitude of mispricing directly but rather sentiment-driven investors' demand for certain types of stocks. If these sentiment investors trade against mean-variance arbitrageurs, the portfolio optimization of the arbitrageurs induces cross-dependencies across expected returns and covariances that can result in equilibrium expected returns that differ from Assumption 1 (for this given $\boldsymbol{X}_t$). To illustrate, consider a CARA-normal model as in Kozak, Nagel, and Santosh (2018) where a measure $(1-\theta)$ of rational arbitrageurs have demand $\frac{1}{a}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t$ and a measure θ of sentiment investors have demand in excess of rational investor demand of $\boldsymbol{X}_t\boldsymbol{d}$ for some vector $\boldsymbol{d}$, i.e., $\frac{1}{a}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t+\boldsymbol{X}_t\boldsymbol{d}$. With total asset supply of one for each asset, collected in vector $\boldsymbol{\iota}$, market clearing implies

$$\boldsymbol{\mu}_t = a\boldsymbol{\Sigma}_t(\boldsymbol{\iota} - \theta\boldsymbol{X}_t\boldsymbol{d}) = \boldsymbol{\Sigma}_t\boldsymbol{X}_t\boldsymbol{\phi}, \tag{44}$$

for some vector $\boldsymbol{\phi}$, where the last equality follows because $\boldsymbol{X}_t$ includes a column of ones. Thus, in this case instead of Assumption 1, we would have

Assumption 2

$$\boldsymbol{\mu}_t = \boldsymbol{\Sigma}_t \boldsymbol{X}_t \boldsymbol{\phi} \tag{45}$$

with some $J \times 1$ *vector* $\boldsymbol{\phi}$.

A closely related assumption appears in Brandt, Santa-Clara, and Valkanov (2009). They assume that mean-variance efficient portfolio weights are linear in characteristics and market portfolio weights, while here Assumption 2 implies that the weights $\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t = \boldsymbol{X}_t\boldsymbol{\phi}$ are linear in characteristics. Similarly, Kozak, Nagel, and Santosh (2020) assume that SDF prices of risk are linear in $\boldsymbol{X}_t$.

The SDF in this case is spanned by GLS factors from GLS cross-sectional regression of $\boldsymbol{z}_{t+1}$ on $\boldsymbol{\Sigma}_t\boldsymbol{X}_t$, or transformations of these factors. We can obtain these factors by replacing $\boldsymbol{X}_t$ in Proposition 1 with $\boldsymbol{\Sigma}_t\boldsymbol{X}_t$ everywhere. We get factors

$$\boldsymbol{f}_{t+1} = \boldsymbol{S}_t'\boldsymbol{X}_t'\boldsymbol{z}_{t+1}, \tag{46}$$

i.e., the GLS factors simplify to univariate factors or transformations thereof (e.g., OLS factors with $\boldsymbol{S}_t = (\boldsymbol{X}_t'\boldsymbol{X}_t)^{-1}$). In other words, one can construct factors that span the SDF solely based on the information in characteristics. No information about $\boldsymbol{\Sigma}_t$ is required to construct these factors! Unfortunately, as we see in the following example that summarizes the univariate factor case, conditional factor means and betas vary over time with $\boldsymbol{\Sigma}_t$, which renders empirical implementation difficult without further assumptions.

Example 7 *Suppose* $\boldsymbol{S}_t = \boldsymbol{I}$. *We then obtain an SDF with factors* $\boldsymbol{f}_{t+1} = \boldsymbol{X}_t'\boldsymbol{z}_{t+1}$. *Factor means and covariances are time-varying,* $\boldsymbol{\mu}_{f,t} = \boldsymbol{X}_t'\boldsymbol{\Sigma}_t\boldsymbol{X}_t\boldsymbol{\phi}$, $\boldsymbol{\Sigma}_{f,t} = \boldsymbol{X}_t'\boldsymbol{\Sigma}_t\boldsymbol{X}_t$, *and factor risk prices are constant:* $\boldsymbol{b}_t = \boldsymbol{\phi}$. *Factor betas* $\boldsymbol{\beta}_t = \left(\boldsymbol{X}_t'\boldsymbol{\Sigma}_t\boldsymbol{X}_t\right)^{-1}\boldsymbol{\Sigma}_t\boldsymbol{X}_t$ *are varying with* $\boldsymbol{\Sigma}_t$.

Under Assumption 2, dimension reduction works in the same way and under the same conditions on the covariance matrix as in Corollary 2.[11]

Whether Assumption 2 or Assumption 1 is more appropriate once a researcher has settled on a specification of $\boldsymbol{X}_t$ is an empirical question. We return to our baseline Assumption 1 for the rest of this section.

IV.B. Conditioning down

For empirical work, our results in terms of conditional moments are not straightforward to work with. In empirical implementation, researchers often like to work with unconditional pricing restrictions and unconditional moments as estimating conditional moments requires additional assumptions about the dynamics of conditional moments.

For this purpose, it is convenient if a model implies that factors' conditional expected returns are constant and either conditional factor betas or factor prices of risk are also constant or depend only on the observable characteristics $\boldsymbol{X}_t$ (and not on $\boldsymbol{\Sigma}_t$). For example, if $\boldsymbol{\beta} = \boldsymbol{X}_t$ and $\boldsymbol{\mu}_{f,t} = \boldsymbol{\mu}_f$, one can implement the factor model in its conditional beta pricing formulation and then condition down to

$$\mathbb{E}[\boldsymbol{z}_{t+1}] = \mathbb{E}[\boldsymbol{X}_t]\boldsymbol{\mu}_f. \tag{47}$$

Alternatively, if $\boldsymbol{b}_t = \boldsymbol{\phi}$ and $\boldsymbol{\mu}_{f,t} = \boldsymbol{\mu}_f$, we have an SDF

$$M_{t+1} = 1 - \boldsymbol{\phi}'(\boldsymbol{z}_{t+1} - \boldsymbol{\mu}_f), \tag{48}$$

which we can rescale to

$$M_{t+1} = 1 - \frac{\boldsymbol{\phi}'}{1 - \boldsymbol{\phi}'\boldsymbol{\mu}_f}\boldsymbol{z}_{t+1} \tag{49}$$

11. In this case, we don't need the additional assumption about expected returns in (35) because its expected returns automatically inherit the lower-dimensional structure through their dependence on the covariance matrix in Assumption 2.

without affecting the pricing implications for excess returns. In this formulation, one can estimate the J constant prices of risk $\boldsymbol{b} = \frac{\boldsymbol{\phi}'}{1-\boldsymbol{\phi}'\boldsymbol{\mu}_f}$ from the J unconditional pricing restrictions $\mathbb{E}[M_{t+1}\boldsymbol{f}_{t+1}] = 0$ without having to model conditional moments.

Recall that our earlier results in Section II expressed factors up to a transformation by a nonsingular matrix $\boldsymbol{S}_t$. We can choose this matrix to generate factors with the desired conditioning-down properties.

Consider first the GLS factors and their transformations. The case we presented in Example 1 with $\boldsymbol{S}_t = \left(\boldsymbol{X}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{X}_t\right)^{-1}$ yields $\boldsymbol{\beta}_t = \boldsymbol{X}_t$ and $\boldsymbol{\mu}_{f,t} = \boldsymbol{\phi}$, so the beta-pricing formulation conditions down nicely, but there is no $\boldsymbol{S}_t$ that produces both prices of risk that do not depend on $\boldsymbol{\Sigma}_t$ and factor means that do not depend on $\boldsymbol{\Sigma}_t$. As a consequence, there does not exist a version of $\boldsymbol{S}_t$ that would yield an SDF that we could estimate without having to model $\boldsymbol{\Sigma}_t$.

Similarly, for OLS factors and their transformations, the case in Example 3 with $\boldsymbol{S}_t = \left(\boldsymbol{X}_t'\boldsymbol{X}_t\right)^{-1}$ yields $\boldsymbol{\beta}_t = \boldsymbol{X}_t$ and $\boldsymbol{\mu}_{f,t} = \boldsymbol{\phi}$, so again the beta-pricing formulation conditions down nicely, but there is no $\boldsymbol{S}_t$ that produces both prices of risk that do not depend on $\boldsymbol{\Sigma}_t$ and factor means that do not depend on $\boldsymbol{\Sigma}_t$.

Under the alternative Assumption 2 about expected returns in Section IV.A, too, there is no specification of $\boldsymbol{S}_t$ that produces, at the same time, prices of risk that do not depend on $\boldsymbol{\Sigma}_t$ and factor means that do not depend on $\boldsymbol{\Sigma}_t$.

IV.C. Orthonormalized characteristics

Empirical work often considers characteristics that are normalized in some fashion. For example, portfolio sorting procedures use only information about cross-sectional ranks of stocks by characteristics, not the value of the characteristics themselves; other methods transform characteristics into cross-sectional ranks and use the rank-transformed characteristics as portfolio weights (Kozak, Nagel, and Santosh 2020); further alternatives include orthonormalizing characteristics such that, after orthonormalization, $\boldsymbol{X}_t'\boldsymbol{X}_t = \boldsymbol{I}$ holds. Common to these methods is that, to varying degrees, they remove time-series variation from characteristics. For example, if the original characteristics matrix includes a column of ones as first column, and characteristics are then orthonormalized using the Gram-Schmidt process, this cross-sectionally demeans all characteristics and removes time-series variation in their cross-sectional variances and correlations.

IV.D. Conditioning down with normalized characteristics

We now show that constructing factors based on such normalized characteristics can be advantageous in light of the requirements we discussed in Section IV.B for unconditional pricing restrictions to imply an SDF with constant factor prices of risk and constant factor means.

However, before we can discuss conditioning down the pricing relationship to unconditional moments, we first need to deal with the fact that if Assumption 1 holds for a given set of original characteristics, it does not necessarily hold for the normalized version of these characteristics. Whether it holds for the normalized version is ultimately an empirical question, but there are plausible reasons to think that it could. To see why, let's focus on the case of orthonormalization and let $\boldsymbol{C}_t$ be the original characteristics matrix and $\boldsymbol{X}_t = \boldsymbol{C}_t \boldsymbol{N}_t^{-1}$ the normalized one, with $\boldsymbol{N}_t = (\boldsymbol{C}_t' \boldsymbol{C}_t)^{\frac{1}{2}}$. What is needed, roughly, is that the normalized characteristics do not contain information about the time-variation in cross-sectional mean, dispersion, or correlation of the original characteristics that the normalization has removed. More precisely, we need that

$$\mathbb{E}[\boldsymbol{N}_t | \boldsymbol{X}_t] = \boldsymbol{N} \tag{50}$$

for some constant matrix $\boldsymbol{N}$. If this holds, and Assumption 1 holds for the original characteristics, i.e., $\mathbb{E}[\boldsymbol{z}_{t+1} | \boldsymbol{C}_t] = \boldsymbol{C}_t \boldsymbol{\phi}_C$, then,

$$\mathbb{E}[\boldsymbol{z}_{t+1} | \boldsymbol{X}_t] = \mathbb{E}\left\{\mathbb{E}[\boldsymbol{z}_{t+1} | \boldsymbol{C}_t] | \boldsymbol{X}_t\right\} = \boldsymbol{X}_t \boldsymbol{\phi}, \qquad \boldsymbol{\phi} = \boldsymbol{N} \boldsymbol{\phi}_C, \tag{51}$$

i.e., we see that the relationship between characteristics and conditional expected excess returns remains linear with constant coefficients $\boldsymbol{\phi}$. In this case, GLS factors constructed based on the normalized characteristics price perfectly all assets conditional on $\boldsymbol{X}_t$. The maximum squared Sharpe ratio attainable conditional on $\boldsymbol{X}_t$ may be lower than conditional on $\boldsymbol{C}_t$, but all of our earlier analysis of the conditions for OLS factors to span the SDF, for factor hedging, and dimension reduction then go through based on the normalized characteristics with conditional moments conditioned on $\boldsymbol{X}_t$.

Normalization of characteristics can be useful if we wish to condition down to unconditional pricing restrictions and obtain an SDF with constant factor prices of risk and constant factor means. For instance, purging characteristics of information about time-varying cross-sectional mean, dispersion, or correlation of characteristics, removes much of the information that in characteristics that could be related to time-variation in $\boldsymbol{\Sigma}_t$. As a consequence, relatively mild assumptions suffice to obtain constant factor prices of risk and constant factor means.

Based on orthonormalized characteristics, the OLS factors in Example 3 have means $\boldsymbol{\mu}_{f,t} = \boldsymbol{\phi}$ and prices of risk $\boldsymbol{b}_t = \boldsymbol{\Psi}_t^{-1}\boldsymbol{\phi}$. So time-variation in $\boldsymbol{\Psi}_t$ is the only remaining source of time-variation in the prices of risk. With orthonormalized characteristics, the assumption that $\boldsymbol{\Psi}_t$ is constant is a relatively weak one. Recall that all conditional moments in our analysis, including $\boldsymbol{\Psi}_t$, are conditioned on $\boldsymbol{X}_t$. Since orthonormalization removes variation over time in the average value of characteristics, their dispersion, and their correlation, there may not be much information left in characteristics that captures time-variation in $\boldsymbol{\Psi}_t$. Therefore, conditional on the normalized characteristics $\boldsymbol{X}_t$, $\boldsymbol{\Psi}_t$ could be constant even it is not constant conditional $\boldsymbol{C}_t$.

For example, consider book-to-market equity ratios. Before normalization, the average book-to-market ratio across firms may have time-series variation that is informative about time-variation in conditional covariances $\boldsymbol{\Psi}_t$. Orthonormalization removes this common variation. Similarly, before normalization, book-to-market ratios may have time-varying cross-sectional dispersion that is informative about time-variation in $\boldsymbol{\Psi}_t$. Orthonormalization removes this information. There could potentially still be some information in, say, the cross-sectional ordering of firms by characteristics each period that could contain information about time-varying in $\boldsymbol{\Psi}_t$, but it seems likely that orthonormalizing removes most of the variation in characteristics that could be informative about time-variation in $\boldsymbol{\Psi}_t$.

If $\boldsymbol{\Psi}_t$ is indeed constant conditional on the orthonormalized characteristics, then prices of risk are constant, $\boldsymbol{b}_t = \boldsymbol{b}$, and hence the SDF

$$M_{t+1} = 1 - \boldsymbol{b}'(\boldsymbol{f}_{t+1} - \boldsymbol{\mu}_f) \tag{52}$$

can be estimated from unconditional pricing restrictions and without estimating a conditional covariance matrix. Thus, orthonormalization combined with a relatively weak assumption about $\boldsymbol{\Psi}_t$ may make it possible to use standard estimation approaches that rely on unconditional moments.

IV.E. Testing

We close this section with a few remarks on testing. The previous analysis made clear that heuristic factor models, such as OLS factors, only span the SDF when the conditional covariance matrix satisfies certain conditions. How can we let the data tell us whether these conditions hold? Going into the sampling theory of estimation and testing is beyond the scope of this paper.[12] Instead, we will highlight population moment conditions that reveal misspecification (and ones that do not). We focus our discussion on OLS factors.

It may seem straightforward to test an OLS factor model. Let $\boldsymbol{f}_{t+1}$ denote the OLS factors from Example 3. In this case we have observable conditional betas $\boldsymbol{\beta}_t = \boldsymbol{X}_t$ and

12. Pezzo, Velu, Zhou, and Wang (2022) build on our population results to develop an asymptotic inference approach based on reduced-rank regression.

constant factor means $\boldsymbol{\mu}_{f,t} = \boldsymbol{\phi}$. Therefore, it may seem natural to simply evaluate whether

$$\mathbb{E}[\boldsymbol{z}_{t+1}] = \mathbb{E}[\boldsymbol{X}_t \boldsymbol{f}_{t+1}] \tag{53}$$

holds in the data. In fact, this is what Fama and French (2020) do in their empirical work when they evaluate an OLS factor model. However, testing the equality (53) just tests whether there is a linear relation between characteristics and expected returns as stated in Assumption 1. If Assumption 1 holds, the equality (53) is true irrespective of whether the conditions in Proposition 2 for OLS factors to span the SDF hold or not. To see this, note that $\mathbb{E}[\boldsymbol{X}_t \boldsymbol{f}_{t+1}] = \mathbb{E}[\boldsymbol{X}_t(\boldsymbol{X}_t'\boldsymbol{X}_t)^{-1}\boldsymbol{X}_t'\boldsymbol{z}_{t+1}] = \mathbb{E}[\boldsymbol{X}_t\boldsymbol{\phi}] = \mathbb{E}[\boldsymbol{z}_{t+1}]$ by Assumption 1. So testing the equality (53) is not a test of the OLS factor asset pricing model.

The key here is that misspecification due to the conditional covariance matrix not satisfying the conditions in Proposition 2 would show up as $\boldsymbol{\beta}_t$ deviating from $\boldsymbol{X}_t$. By assuming $\boldsymbol{\beta}_t = \boldsymbol{X}_t$, the approach of Fama and French (2020) assumes away any misspecification of the SDF.

One way to testing for misspecification is to construct hedged factors as in Sections II.C and II.D. If the hedged factors achieve a higher Sharpe ratio than the OLS factors, the OLS factors do not span the SDF. We implement this approach empirically in the next section.

V. Empirical analysis

Our analysis so far provides conditions on the conditional covariance matrix of individual stock returns under which OLS factors (and transformations thereof) span the SDF, as well as conditions under which dimension-reduction via principal components analysis of OLS factor portfolios yields the same factors as IPCA. Do these conditions hold empirically for various combinations of characteristics-based factors used in the prior literature?

Directly answering this question by comparing the maximum squared Sharpe ratio attainable with OLS factors to the maximum squared Sharpe ratio of GLS factors is difficult because constructing GLS factors requires the estimation and inversion of a large conditional covariance matrix for an unbalanced panel of thousands of stocks. Below we employ a heuristic approach for estimating this covariance matrix, but we also use our earlier results on iterated factor hedging to shed more light on this issue. The logic behind this latter approach is that if a set of OLS factors does not span the SDF, then hedging the factors should improve the maximum squared Sharpe ratio. If a set of OLS factors already spans the SDF, then factor hedging should not yield an improvement. In fact, empirically, with estimated moments that are contaminated with estimation error, factor hedging might lead to a deterioration in the Sharpe ratio.

V.A. Data and factor construction

We use rank-transformed standardized stock characteristics from Kozak (2019) and daily stock returns from July 1972 to December 2021. We apply several filters to preserve characteristics with maximum data availability. In particular, we remove any characteristics for which more than 25% of the observations in the panel of firms are missing. We remove any time periods in the early part of the sample for which less than 500 firms are available. We also remove firms whose past market caps do not exceed 0.0025% of the aggregate stock market capitalization (e.g., firms with market capitalizations less than \$1 billion on a \$40 trillion aggregate stock market valuation). Lastly, we fill in any missing characteristics with their cross-sectional means, which are equal to zero for standardized data.[13] We collect the resulting 34 rank-transformed standardized characteristics, including the unitary characteristic, for each of the stocks in the monthly characteristics matrix $\boldsymbol{X}_t$.[14] Our final dataset contains 594 months of monthly characteristics and daily returns on 9,201 stocks.

As we discussed in Section IV.D, normalizations such as rank-transformation remove time-varying components of characteristics. Unlike orthonormalization, rank-transformation

13. We also use a dataset with no market capitalization filters, a dataset with imputed characteristic values using a more advanced imputation method, as well as other datasets based on different and broader sets of characteristics (see Appendix, Section B).

14. Table I provides the list of characteristics we use.

does not remove information about time-varying correlations, but time-varying components of cross-sectional means and dispersion of characteristics are removed. On one hand, removing these components may restrict the investment opportunity set and lower the maximum squared Sharpe ratio that is attainable. On the other hand, the conditions necessary for means, covariances, and risk prices of OLS factors to be constant are more likely to hold. If these moments are constant, the unconditional maximum squared Sharpe ratio of the factors is equal to its (constant) conditional version and we can evaluate factor models based on the unconditional maximum squared Sharpe ratio that the factors attain.

We consider several types of factor constructions in line with our theoretic developments in the earlier part of the paper: (i) *univariate factors* with weights given by $\boldsymbol{X}_t'$; (ii) *orthonormalized factors* with orthonormalized weights $(\boldsymbol{X}_t'\boldsymbol{X}_t)^{-\frac{1}{2}}\boldsymbol{X}_t'$ constructed using the singular value decomposition;[15] (iii) *OLS factors* with weights $(\boldsymbol{X}_t'\boldsymbol{X}_t)^{-1}\boldsymbol{X}_t'$; and (iv) *GLS factors* with weights equal to $(\boldsymbol{X}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{X}_t)^{-1}\boldsymbol{X}_t'\boldsymbol{\Sigma}_t^{-1}$.

We update factor weights at the end of each month t. To avoid intra-month trading, we evaluate all return-based performance metrics using stock returns aggregated to monthly frequency, which corresponds to monthly buy-and-hold returns. For factor hedging and cross-sectional regressions that rely on rolling covariance estimates, we use the most recently available data up to that point in time (i.e., up until the end of a prior month).

Our main analysis is conducted in full sample. We also report out-of-sample results using a split-sample approach. Specifically, we split the sample into two parts: pre-2005 and 2005–present. When reporting out-of-sample maximum squared Sharpe ratios, we use the sample covariance matrix of monthly factor returns as an estimate of the unconditional factor covariance matrix and factor means as estimates of unconditional expected excess returns on

15. Let the singular value decomposition of $\boldsymbol{X}$ be given by $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}'$. Then orthonormalized characteristics $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-\frac{1}{2}}$ can be constructed simply as $\boldsymbol{U}\boldsymbol{V}'$. This orthogonalization is also known as Löwdin symmetric orthogonalization, as well as Mahalanobis whitening or ZCA (zero-phase component analysis; see Murphy (2023)). Unlike Gram-Schmidt orthogonalization, it treats all characteristics symmetrically and has an appealing property that orthonormalized columns are the least distant from the original columns of $\boldsymbol{X}_t$ (in the least squares sense). That is, it indicates the gentlest pushing of each characteristic in the cross-section in order to get them to be orthogonal. Such orthogonalization, therefore, partially preserves the economic identity of characteristics and factors which is a useful property in our setting.

the factors, with both estimated in the pre-2005 sample of returns. Combining information from covariances and means, we compute MVE portfolio weights which we then fix and apply to the 2005–present sample of monthly stock returns. We then compute annualized unconditional squared Sharpe ratios of these series in the out-of-sample period.

V.B. GLS factors

While GLS factors defined in Proposition 1 are mean-variance efficient, they are generally infeasible because their construction requires knowledge of the conditional covariance matrix of individual stock returns, $\boldsymbol{\Sigma}_t$. We propose a heuristic non-parametric approach for estimating $\boldsymbol{\Sigma}_t$.

We compute rolling covariances in 3-year windows of daily individual stock returns (up until the end of previous month), we use PCA to extract 30 factors, and we approximate the conditional covariance matrix based on this 30-factor model at every time t. We set idiosyncratic variances equal to the cross-sectional mean of idiosyncratic stock-level variances for every time t.[16]

This approach is conceptually similar to what factor hedging is attempting to achieve: extract information from the empirical covariance matrix. Hedging does not estimate the full covariance matrix, however, while this method does.

V.C. Hedging factors

As an alternative to GLS-based approaches which rely on the estimate of the rolling or highly parametric covariance matrix of stock returns, we implement the factor hedging procedure of Section II.C. We view this procedure as an iterative approach of extracting information from the inverse of the covariance matrix of stock returns without the need to fully estimate this matrix.

16. Our rolling covariance matrix approach potentially uses some info not in $\boldsymbol{X}_t$, which is a slight deviation from our theoretical setup.

We compute daily rolling covariances of individual stocks returns with the factors within overlapping backward-looking 3-year windows. We then regress these daily covariances on the characteristics $\boldsymbol{X}_t$. The residuals from these regressions give us daily portfolio weights $\boldsymbol{W}_{h,t}$ of the hedging portfolios which we then use to calculate daily hedging factor returns. This completes the first step in the approach we outlined in Section II.C.

For our main analysis, the second and third steps follow the procedure we outlined in the main text in Section II.C: we calculate stocks' covariances with the hedge portfolio returns so that we can modify stocks' weights in the factor portfolios to remove unpriced risks, and then regress the factor portfolio weights on these covariances to obtain residual factor portfolio weights that have been purged of unpriced risk exposure. We define characteristics associated with these factors to be "hedged characteristics." To construct iterated hedged factors, we repeat this procedure multiple times.[17]

V.D. Empirical performance of hedged factors

Figure I shows improvement in average in-sample MVE portfolio's squared Sharpe ratios constructed from hedged OLS factors relative to unhedged factors, in %. We run the hedging procedure for up to three rounds of hedging. We calculate these improvements for OLS factor models with different numbers of characteristic-based factors from one to fifteen, in addition to the constant characteristic which is implicitly included in all models. Since there are different possible subsets of J factors from the full 34 OLS factors, we draw, for each J, 10,000 random subsets of J factors. Figure I shows the percentage improvement in the maximum squared Sharpe ratio averaged across these random subsets for each J.

As the figure shows, the benefit of hedging decreases as the number of characteristics

17. In addition to this approach we also implement the approach of DMRS that we discussed in footnote 4 of Section II.C. That is, in the second and third steps we purge the ad-hoc factors from unpriced risks by regressing the daily univariate factors on the daily hedge portfolio returns. The parameters of this regression are estimated using full sample and then used to construct residuals (as an alternative, we also implemented and tested estimating the parameters of this regression in rolling or expanding windows). The residuals are the hedged factors. We refer to this type of hedging as "DMRS hedging." The results are reported in Appendix Tables A.IV, A.V and A.VI.

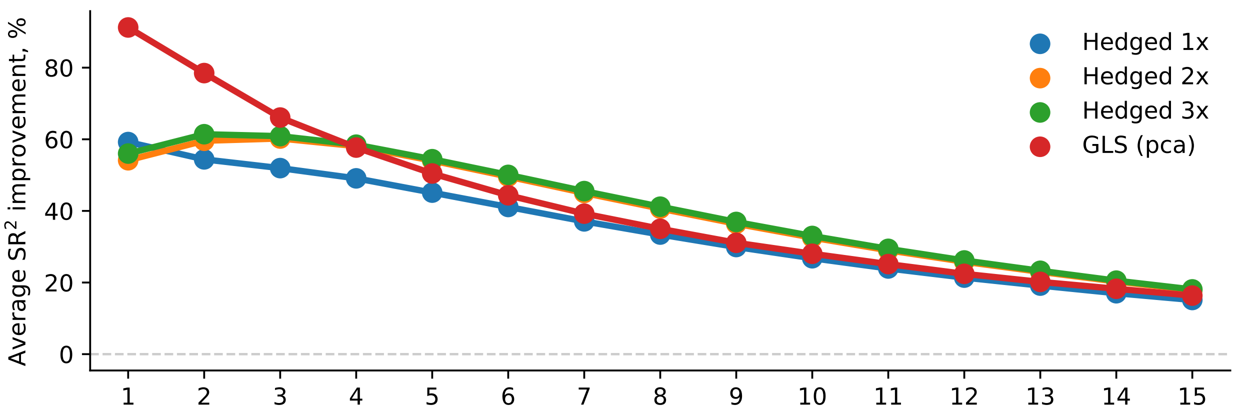

FIGURE I: **In-sample maximum squared Sharpe ratio improvement due to iterative hedging of OLS factors.** The plot shows improvement, in %, of annualized average in-sample maximum squared Sharpe ratio of hedged OLS factors relative to unhedged factors, for all models with a constant and 1–15 additional factors. We hedge the factors up to three times. We also report performance of the GLS factors which use the sample conditional covariance matrix of individual stock returns estimated using the rolling PCA procedure outlined in Section V.B. For each number of factors on the x-axis, results are averaged across 10,000 models with this number of factors randomly drawn from the set of all factors.

increases. This is what we anticipated in our discussion of Proposition 2. Including a large number of characteristics makes it more likely that loadings on major sources of covariances are spanned by the columns of $\boldsymbol{X}_t$. This renders violations of the conditions of Proposition 2 quantitatively less important. As a consequence, a large number of OLS factors approximately spans the SDF and factor hedging provides little additional benefit.

The benefit trends towards zero and might even turn negative when J is large. Under population moments, as in our earlier theoretical analysis, hedging would never lead to a deterioration of the Sharpe ratio in sample. However, with estimated moments, estimation error contaminates the hedging procedure and hedging can then lead to a deterioration, especially out of sample.

The figure also shows that there can be a benefit from iterating on the hedging procedure using the iterated hedging approach that we developed in our theoretical analysis. This benefit is larger if the number of factors is relatively small. For example, with $J = 2..5$, hedging a single time leads to an improvement in average maximum squared Sharpe ratio

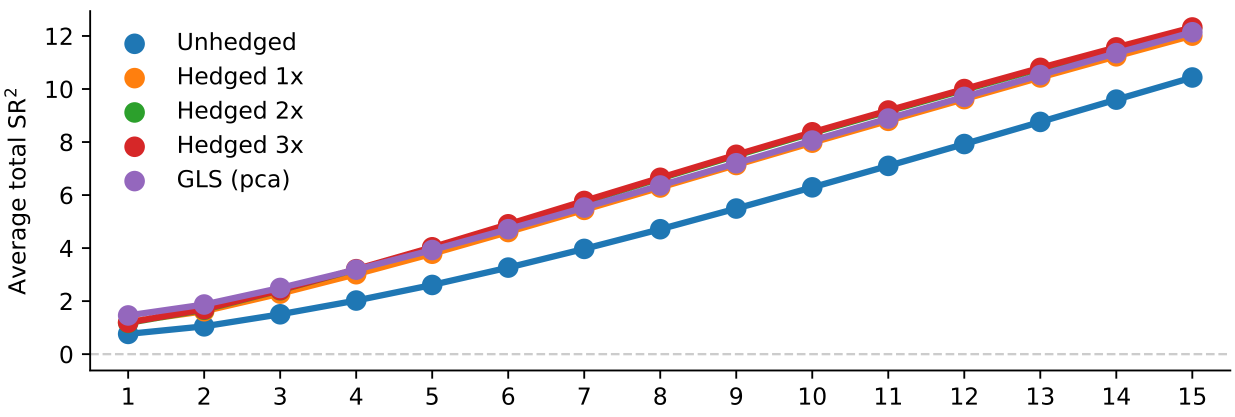

FIGURE II: **In-sample maximum squared Sharpe ratio of hedged OLS factors.** The plot shows annualized average in-sample maximum squared Sharpe ratios of unhedged and hedged OLS factors, as well as GLS factors. We hedge the factors up to three times. The latter use the sample conditional covariance matrix of individual stock returns estimated using the rolling PCA procedure outlined in Section V.B. For each number of factors on the x-axis, results are averaged across 10,000 models with this number of factors randomly drawn from the set of all factors.

of about 50%. Hedging one more round raises this number to about 60%, on average. The marginal benefit of each additional round of hedging is small. For $J \geq 15$ the benefit of the second round of hedging largely dissipates.

The red line in Figure I depicts average squared Sharpe ratio improvements generated by GLS factors based on a conditional covariance matrix of individual stock returns estimated using the rolling PCA procedure outlined in Section V.B. Recall that iterative hedging is essentially a way of using information from this (inverse) covariance matrix without the need to estimate the entire matrix. As such, we would expect hedging to perform similarly to GLS. This is exactly what we see in the figure: GLS factors generate improvements in maximum squared Sharpe ratios of broadly similar magnitude—slightly higher than hedging for models with one or two factors, and about the same as hedged factors for models with three or more factors. Overall, this result suggests that factor hedging performs as intended. It is a useful simpler alternative to GLS that avoids estimation of the full conditional covariance matrix of individual stock returns.

Figure II demonstrates the effect of these improvements on the total average squared Sharpe ratio, in levels, for models with unhedged, hedged (up to three times), and GLS factors. Similar to Figure I, we see that using hedged or GLS factors leads to sizeable improvements in maximum squared Sharpe ratios. The level of improvement decays only slightly as J increases in the range 1..15, indicating that although the relative benefit of hedging decays with J, as we have seen in Figure I, hedging still leads to roughly the same level increases in squared Sharpe ratio (of around 2.0) in our dataset for $J \leq 15$. For larger J, the level of improvement continues to decay, however, and approaches zero when all factors are included (see Table II below).

In addition to studying the OLS factors, we also repeat the above analysis for univariate and orthonormalized factors in Appendix, Section A. We find that average squared Sharpe ratio improvements of univariate factors decay much more slowly with the number of factors J, and that there is higher benefit to hedging more than one round. These results suggest that univariate factors might be more contaminated with unpriced risks than OLS factors and there is more room for correcting these inefficiencies with hedging or GLS factor constructions, even for models with a large number of factors. Results for orthonormalized factors are in between those for OLS and univariate factors. We summarize all these results in Appendix Table A.I.

We now turn our attention to the edge cases of the Figure I. First, we look at models with two factors: the level factor corresponding to the constant vector in $\boldsymbol{X}_t$ and one additional factor based on one characteristic at a time. Second, we look at models that use all available factors. We can consider each of these models individually without having to rely on random sampling as we have done previously.

Table I shows how hedging changes the in-sample squared Sharpe ratio of various specific two-factor models. The first column shows the results for unhedged factors, the next three columns hedge factors iteratively up to three times, and the last column employs GLS hedging. In each row of the table, the characteristics matrix $\boldsymbol{X}_t$ includes a constant and the

TABLE I: In-sample maximum squared Sharpe ratios of two-factor OLS models.
We report in-sample maximum annualized squared Sharpe ratios of all models which use OLS factors (first column), OLS hedged factors for $n = 1..3$ rounds, as well as approximate GLS factors (the last column). All models include two characteristics in $\boldsymbol{X}_t$: a constant, and one of the characteristics listed in the rows. GLS factors use a non-parametric covariance matrix estimated via PCA applied to 3-year rolling windows of daily stocks returns. The row labeled "ER" uses fitted values from a panel regression of returns on all characteristics as a standalone characteristic. The last row averages the numbers across all models. * indicate $p < 0.05$ of a one-sided test of the squared Sharpe ratio difference of the given model relative to the unhedged benchmark (first column). ** indicate $p < 0.01$.

	OLS	Hedged n times			GLS
		1	2	3	
Size	0.37	0.73*	0.56	0.59	0.80**
Value (A)	0.76	1.19*	1.23*	1.22*	1.21**
Gross Profitability	0.52	0.97**	0.91*	0.86*	1.13**
F-score	0.83	1.63**	1.86**	1.86**	2.18**
Debt Issuance	0.43	0.76*	0.75	0.77*	1.00**
Share Repurchases	0.71	1.35**	1.40*	1.54**	1.39**
Net Issuance (A)	1.14	1.94**	2.03**	2.17**	1.88**
Asset Growth	0.86	1.23*	1.29*	1.35*	1.48**
Asset Turnover	0.58	0.79	0.60	0.65	0.90*
Gross Margins	0.41	1.00**	0.85*	0.94**	1.25**
Earnings/Price	0.71	1.22*	1.10	1.09	1.30**
Investment/Capital	0.60	0.94*	0.89	0.91	1.07**
Investment Growth	0.82	1.16*	1.22*	1.23*	1.47**
Sales Growth	0.73	0.95	0.85	0.93	1.06*
Leverage	0.53	0.75	0.76	0.72	0.87*
Return on Assets (A)	0.44	0.84*	0.71	0.73*	1.07**
Return on Book Equity (A)	0.44	0.81*	0.66	0.70	1.01**
Sales/Price	0.65	0.87	0.79	0.81	0.93*
Momentum (6m)	0.36	0.65*	0.66*	0.65*	0.92**
Industry Momentum	1.04	1.62*	1.55*	1.52*	1.97**
Momentum (12m)	0.72	1.13*	1.13*	1.10*	1.56**
Momentum-Reversals	0.46	0.81*	0.76	0.72	0.87**
Value (M)	0.60	0.90*	0.94*	0.94*	1.11**
Net Issuance (M)	1.11	1.87**	1.86*	1.89*	2.35**
Short-Term Reversals	0.73	1.62**	1.64**	1.64**	2.12**
Idiosyncratic Volatility	0.75	1.15	0.96	0.91	1.49**
Beta Arbitrage	1.09	0.80	0.88	0.91	1.28
Industry Rel. Reversals	1.50	2.60**	2.63**	2.66**	3.37**
Price	0.37	0.67*	0.69*	0.67*	0.86**
Firm's age	0.55	1.09*	0.98*	0.95*	1.11**
Share Volume	0.83	0.83	0.78	0.77	1.26**
Exchange Switch	0.83	1.42**	1.36**	1.35**	1.42**
IPO	0.43	0.84*	0.72	0.73	0.88**
ER	7.05	9.13*	9.63**	9.73**	11.32**
Average	0.88	1.36	1.34	1.36	1.64

characteristic listed in this row. For all of these, since only one characteristics is used, it is highly unlikely that the conditions hold that are required by Proposition 2 for OLS factors to span the SDF. Hedging the factors should therefore improve the Sharpe ratio. Consistent with this logic, we find improvements from hedging for every characteristic, and the gain is often substantial.

To interpret this correctly, it is important to keep in mind that the failure of the unhedged factors to span the SDF is not a simple consequence of the fact that two-factor models omit other characteristics that are informative about expected returns but are left out from the two-factor model. The hedged factors do not use any information from these other characteristics either. Instead, the reason for the inferiority of the unhedged factors is that a single characteristic is not enough to satisfy the conditions in Proposition 2 for OLS factors to span the SDF that prices assets conditional on this single characteristic.

To see this more clearly, we report squared Sharpe ratios based on approximate GLS factors in the last column of the table. We use a non-parametric covariance matrix estimated via PCA applied to 3-year rolling windows of daily stocks returns to estimate GLS factors. The table shows that hedging OLS factors moves their squared Sharpe ratios in the direction of the GLS factors. GLS factors achieve higher in-sample squared Sharpe ratios than hedged factors do.

As the table shows, there is considerable heterogeneity in how much hedging or GLS adjustments improve the Sharpe ratio. Characteristics like short-term reversals, net issuance, gross margins show dramatic improvements of more than 100% with three rounds of hedging, and even more so when GLS factors are considered, while others show little in-sample improvement. The bottom row shows that on average, across all portfolios, maximum squared Sharpe ratios increase from 0.88 (unhedged factors) to 1.36 (after three rounds of hedging), to 1.64 for GLS factors. These improvements in squared Sharpe ratios are statistically significant. To demonstrate this we conduct a one-sided 5% or 1% test based on Barillas, Kan, Robotti, and Shanken (2020) which compares the squared Sharpe ratio of the model in question to the benchmark OLS model with no hedging (the first column). We use * and ** to indicate significance at these two levels, respectively.

Lastly, we construct a composite characteristic which uses fitted values from a panel regression of returns on all characteristics (row labeled "ER"). This characteristic summarizes expected return predictability of all original characteristics, but uses a single factor and thus generally does not satisfy the conditions in Proposition 2 for OLS factors to span the SDF that prices assets conditional on this single composite characteristic. As such, it is a natural candidate for hedging or GLS factor construction. The table shows that benefits of hedging for this characteristics are substantial. Hedging three rounds achieves a squared Sharpe ratio of 9.63 (from 7.05), while GLS constructions rise this number as high as 11.32.[18] These results indicate that the conditions in Proposition 2 are indeed likely to be violated. In other words, sorting stocks on fitted expected returns preserves information in means but largely discards information in covariances, which prevents the factor from reaching mean-variance efficiency.

Appendix Table A.III shows the out-of-sample results. Overall, they are consistent with in-sample results. Hedging once increases the squared Sharpe ratios marginally, from 0.47 to 0.61 on average across all models with no out-of-sample benefit to hedging more rounds. GLS factors achieve the squared Sharpe of 0.89. Hedging the "ER" characteristic also improves squared Sharpe out of sample, to 1.61 (from 1.31) after three rounds of hedging, and as high as 2.56 when using GLS factors. Table A.II reports the out-of-sample results for univariate factors and factors based on orthonormalized characteristics. Out-of-sample results exhibit similar patterns as the in-sample results, but the magnitude of effects is diminished.

In Table II, we consider the effect of hedging on the models with the full set of 34 factors, both in sample (top panel) and out of sample (bottom panel). The first column shows the maximum squared Sharpe ratios of the original unhedged models, while the following five

18. These GLS constructions can be interpreted as direct estimates of an MVE portfolio constructed from individual stock returns in (4), where stock-level expected returns $\boldsymbol{\mu}_t$ are estimated via a panel regression of returns on all characteristics, and the stock-level covariance matrix of returns $\boldsymbol{\Sigma}_t$ is estimated using the non-parametric PCA-based approach discussed above.

TABLE II: Maximum squared Sharpe ratios of hedged factors.

We report in-sample (top panel) and out-of-sample (bottom panel) annualized maximum squared Sharpe ratio of the MVE portfolio constructed from 34 unhedged (first column) or hedged up to five times factors. Rows correspond to three types of factors we discuss in Section V.A. * indicate $p < 0.05$ of a one-sided Barillas, Kan, Robotti, and Shanken (2020) test of the squared Sharpe ratio difference of a given model relative to the unhedged benchmark (first column). ** indicate $p < 0.01$.

	Unhedged	Hedged n times				
		1	2	3	4	5
				In-sample		
Univariate	13.8	16.4*	17.2**	17.3**	17.4**	17.4**
Orthonormal	18.0	20.1*	20.3*	20.4*	20.5*	20.5*
OLS	21.3	21.8	21.5	21.6	21.6	21.6
				Out-of-sample		
Univariate	1.3	1.5	2.1	2.3	2.4	2.5
Orthonormal	3.4	3.1	3.4	3.6	3.6	3.7
OLS	4.0	3.8	3.9	4.1	4.1	4.2

columns hedge the factors up to five times. We use * and ** to indicate significance of the squared Sharpe ratio difference at the 5% and 1% level, respectively, using the one-sided test of Barillas, Kan, Robotti, and Shanken (2020).

The results in the table are consistent with our previous findings and intuition. First, hedging raises the squared Sharpe ratio for univariate factors (13.8 to 17.4 in sample and 1.3 to 2.5 out of sample). For orthonormal and OLS factors the increases are small, consistent with our previous findings that the benefits of hedging decay as the number of factors increases.

To summarize, the results in this section demonstrate sizeable benefits of hedging and GLS factor constructions in terms of squared Sharpe ratio improvements. Hedging is especially beneficial for univariate factors. The benefits of hedging for OLS factors diminish quickly as the number of factors increases because the large number of characteristics renders violations of the conditions of Proposition 2 quantitatively less important. As a consequence, when the number of OLS factors is large, they approximately span the SDF and factor hedging provides little additional benefit. In the same spirit, hedging GLS factors does not lead to any increase in the squared Sharpe ratio since these factors are already approximately mean-variance efficient.

V.E. Dimensionality reduction

Our final empirical analysis looks at dimensionality reduction. In Section III, we showed the conditions necessary for dimensionality reduction to be possible. We also showed a few ways how to proceed with dimensionality reduction and how these approaches are related. In this section we explore and compare these methods empirically.

In particular, in the discussion of Example 5 we showed that the factors $\boldsymbol{f}_{IPCA,t+1}$ in (39) satisfy a conditional version of the first-order conditions in Kelly, Pruitt, and Su (2019) that define the IPCA estimator. This means that a conditional equivalent of an IPCA estimator can be constructed using PCA on managed portfolios, even in the case where characteristics are not orthonormalized, if the managed portfolios are constructed as OLS factors. In particular, applying PCA to OLS portfolios recovers $\boldsymbol{Q}_t$. By applying this matrix to univariate portfolios $\boldsymbol{X}_t'\boldsymbol{z}_{t+1}$ and further transforming them by $(\boldsymbol{Q}_t'\boldsymbol{X}_t'\boldsymbol{X}_t\boldsymbol{Q}_t)^{-1}$ as in (39) yields our version of the IPCA estimator. The assumption of time-constant $\boldsymbol{Q}_t$ and $\boldsymbol{\Lambda}_t$ can justify working with a constant $\boldsymbol{Q}$ extracted from an average conditional, or approximately unconditional, covariance matrix.[19]

Similarly, Example 6 showed that under Assumption (35) we should apply PCA to univariate portfolios constructed using orthonormalized characteristics to obtain PPCA factors from Kim, Korajczyk, and Neuhierl (2021).

As benchmarks for comparison, we also apply PCA to univariate portfolios as motivated by Kozak, Nagel, and Santosh (2018) and Kozak, Nagel, and Santosh (2020) (we denote this specifications a "SCS"). Lastly, we include a GLS analogue of IPCA, labeled as "IPCA (GLS)", that constructs the factors as $\boldsymbol{f}_{t+1} = \left(\boldsymbol{Q}_t'\boldsymbol{X}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{X}_t\boldsymbol{Q}_t\right)^{-1}\boldsymbol{Q}_t'\boldsymbol{X}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{z}_{t+1}$, where $\boldsymbol{Q}_t$

19. In practice, however, the theoretic equivalence between our analytic IPCA approach and the iterative procedure of Kelly, Pruitt, and Su (2019) might not hold exactly if this time-constancy assumption is violated, or if the assumptions in Corollary 2 about the covariance matrix do not hold and hence the dimension-reduction to K factors approach is misspecified.

TABLE III: Dimensionality reduction: Comparing different portfolio-formation approaches

The table reports in-sample (top panel) and out-of-sample (bottom panel) maximum annualized squared Sharpe ratios of N PCs (columns) of factors from one of four portfolio-formation approaches (rows): (i) univariate from Kozak, Nagel, and Santosh (2018) and Kozak, Nagel, and Santosh (2020) (SCS), (ii) IPCA from Kelly, Pruitt, and Su (2019) implemented as in Example 5, (iii) PPCA from Kim, Korajczyk, and Neuhierl (2021) implemented as in Example 6, and (iv) GLS analogue of IPCA factors as explained in the text. Out-of-sample results are based on a split sample estimation before/after 2005.

	1	2	3	4	5	6	7	8	9	10	11	12
	In-sample											
SCS	0.2	0.6	0.9	1.2	3.1	3.1	3.1	4.4	4.7	4.7	7.9	8.1
IPCA	0.3	1.3	4.1	4.5	7.0	7.7	11.9	12.6	13.7	14.4	14.8	15.3
PPCA	0.3	0.3	0.7	2.5	8.3	8.3	8.7	12.0	12.0	13.2	13.2	13.3
IPCA (GLS)	0.6	1.3	11.1	10.9	12.0	12.9	16.4	16.8	16.7	16.5	16.3	16.4
	Out-of-sample											
SCS	0.1	0.2	0.4	0.5	0.4	0.3	0.3	0.6	0.8	0.8	1.6	1.5
IPCA	0.3	0.1	0.7	0.8	1.0	1.1	2.1	2.2	2.6	3.1	3.5	3.8
PPCA	0.2	0.2	0.4	1.0	1.6	1.3	1.2	3.0	2.4	3.2	3.1	3.1
IPCA (GLS)	0.4	0.2	2.7	2.2	2.8	2.8	4.7	4.8	3.8	3.8	3.8	3.7

are eigenvectors from PCA of the GLS factors. For all approaches, we apply PCA to monthly returns.

We now compare empirical performance of these methods of dimensionality reduction in terms of unconditional mean-variance efficiency. Table III reports in-sample and out-of-sample maximum annualized squared Sharpe ratios of these extracted latent factors for each of the four portfolio-formation approaches. We report our results by varying the number of latent factors from 1 to 12 (shown in columns). To compute out-of-sample metrics we split the sample in 2005, estimate mean-variance optimal factor combination in the earlier part of the sample using daily returns, and compute squared Sharpe ratios in the latter part using these pre-2005 weights and monthly returns.

The table shows that our analytical versions of IPCA factors from Example 5 and PPCA factors from Example 6 perform better than PCA on simple univariate factors (SCS). The primary reason for this improvement is the additional linear transformation step in the IPCA procedure. This result is similar to our previous finding that OLS factors perform better than univariate factors in terms of being less contaminated by non-priced risks. Performing an OLS transformation on PCA-implied "characteristics" delivers the same benefit.

Equation (41) shows that PPCA can be thought as a simple PCA on univariate portfolios as in the SCS approach, but applied to orthonormalized characteristics. That is, PPCA uses *only* information from orthonormalized characteristics and disregards the information from the original characteristics. As discussed in Section IV.D normalization of characteristics removes time-series variation in their cross-sectional variances and correlations, but can be advantageous for conditioning down the models. The maximum squared Sharpe ratio attainable conditional on orthonormalized characteristics might therefore be lower than that of the original characteristics. Table III shows that Sharpe ratio deterioration is small in the data: squared Sharpe ratios attainable from orthonormalized characteristics (PPCA) are roughly the same as the ones from the IPCA method but significantly higher than the ones attainable from SCS factors.

Note that if we work with cross-sectionally orthonormalized characteristics directly, all methods discussed above become equivalent. This is because in this case $\boldsymbol{X}_t'\boldsymbol{X}_t = \boldsymbol{I}$ so any OLS transformations drop out. That is, other methods, such as IPCA, become equivalent to PPCA if we restrict their information set to orthonormalized characteristics. Without this restriction, these other methods, in principle, use a broader information set and could outperform PPCA. However, in practice, we find that the difference in performance is small.

The last row in each panel focuses on the GLS analogue of IPCA factors. We see that these factor models achieve squared Sharpe ratio improvements with fewer factors than their counterparts that ignore information in the covariance matrix of stock returns. The GLS version of IPCA achieves the highest squared Sharpe ratios and only needs 7-8 factors to get there.

The fact that the GLS version of IPCA performs better than IPCA when the number of latent factors is low suggests that latent factor models with a small number of factors can potentially benefit from hedging. We investigate this conjecture in Table A.VII in the Appendix. We find that hedging can indeed improve the performance of latent factor models, especially the ones that are not OLS transformed, such as the SCS model. Table A.VIII in the Appendix shows that the same is true in out-of-sample data.

In summary, we find that latent factor models perform quite differently depending on how their factors are constructed. In general, OLS-transformed characteristics lead to more efficient factors with less contamination from unpriced risks, both for simple and latent factors. If we restrict the information set to include only orthonormalized characteristics, which is what PPCA does, all methods become equivalent under this information set and perform on par with the OLS-transformed factors. We see some benefits of hedging or GLS adjustments for latent factors, especially the ones that are not OLS-transformed.

VI. Conclusion

Heuristic factor construction by sorting on firm characteristics, weighting by characteristics, or computing OLS cross-sectional regression slopes does not use information about the covariance matrix of individual stock returns. As a consequence, these heuristic factors span the SDF that prices individual stocks only if the covariance matrix satisfies certain special conditions. We work out what these conditions are and obtain a number of insights.

First, horse races between direct prediction of excess returns with characteristics and heuristic characteristics-based factor models, or between different heuristic factor models, have no economic content other than exposing the shortcomings of heuristic factor construction that neglects covariance matrix information. Results from such horse races do not lead to insights about competing economic theories of risk premia and mispricing.

Second, when the individual stock return covariance matrix satisfies conditions such that OLS cross-sectional regression slope factors span the SDF, then nonsingular transformations of OLS factors span the SDF, too, including univariate factors in which stocks in each factor are weighted by a single characteristic. Choice among these different transformations is then a matter of convenience, for example, to obtain suitable conditioning-down properties. Empirically, these conditions do not hold exactly, and OLS factors seem to generally get closer to spanning the SDF.

Third, the conditions on the covariance matrix that allow OLS factors, or transformations thereof, to span the SDF are more likely to hold when the number of characteristics employed by the econometrician is larger. Additional characteristics can help even if they are unrelated to expected returns as long as they help to capture important sources of stock return covariances. We find empirical support for this prediction.

Fourth, heuristic factor models that employ only a small number of characteristics can benefit from purging unpriced risks using hedging methods. Compared with unhedged factors, hedged factors can span the SDF under weaker conditions on the covariance matrix of individual stock returns. Hedging unpriced risks effectively incorporates some information about the covariance matrix into factor construction, but without requiring inversion of a large covariance matrix. Consistent with our theoretical results, we find that hedging benefits are largest for small-scale factor models while OLS factor models with a large number of factors are already close to spanning the SDF.

Fifth, iterating on these hedging procedures allows further relaxation of the conditions on the covariance matrix. Empirically, we find modest benefits from iterated hedging for small-scale factor models, but the benefits from iteration are small for models with a large number of factors.

Sixth, when the relationship of expected returns and covariance matrix to characteristics has a lower-dimensional structure such that information in J characteristics can be captured by $K < J$ characteristics, then the SDF can be spanned by K factors without requiring inversion of a large covariance matrix. Under the conditions on the covariance matrix that allow the factors to span the SDF, simple PCA on OLS factors is equivalent the IPCA method of Kelly, Pruitt, and Su (2019), and simple PCA on univariate factors constructed from orthonormalized characteristics is equivalent to the PPCA method of Kim, Korajczyk, and Neuhierl (2021).

Overall, our results provide the conceptual foundations for the construction, hedging, and dimension-reduction of reduced-form characteristics-based factors that was missing so far in the vast empirical literature on factor models in cross-sectional asset pricing.

References

Back, Kerry, Nishad Kapadia, and Barbara Ostdiek, 2015, "Testing factor models on characteristic and covariance pure plays," Working paper, SSRN.

Barillas, Francisco, Raymond Kan, Cesare Robotti, and Jay Shanken, 2020, "Model comparison with sharpe ratios," *Journal of Financial and Quantitative Analysis* 55, 1840–1874.

Brandt, Michael W, Pedro Santa-Clara, and Rossen Valkanov, 2009, "Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns," *Review of Financial Studies* 22, 3411–3447.

Chib, Siddhartha, Yi Chun Lin, Kuntara Pukthuanthong, and Xiaming Zeng, 2021, "Slope Factors Outperform: Evidence from a Large Comparative Study," Working paper, Washington University.

Daniel, Kent, Lira Mota, Simon Rottke, and Tano Santos, 2020, "The Cross-Section of Risk and Return," *Review of Financial Studies* 33, 1927–1979.

Daniel, Kent, and Sheridan Titman, 1997, "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns," *Journal of Finance* 52, 1–33.

Davis, James, Eugene F. Fama, and Kenneth R. French, 2000, "Characteristics, Covariances, and Average Returns: 1929 to 1997," *Journal of Finance* 55, 389–406.

Fama, Eugene F., 1976, "Foundations of Finance."

Fama, Eugene F., and Kenneth R. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics* 33, 23–49.

Fama, Eugene F, and Kenneth R French, 2015, "A Five-Factor Asset Pricing Model," *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F, and Kenneth R French, 2020, "Comparing Cross-Section and Time-Series Factor Models," *Review of Financial Studies* 33, 1891–1926.

Gerakos, Joseph, and Juhani T Linnainmaa, 2018, "Decomposing Value," *Review of Financial Studies* 31, 1825–1854.

Giglio, Stefano, and Dacheng Xiu, 2021, "Asset Pricing with Omitted Factors," *Journal of Political Economy* 129, 1947–1990.

Grinblatt, Mark, and Konark Saxena, 2018, "When Factors Do Not Span Their Basis Portfolios," *Journal of Financial and Quantitative Analysis* 53, 2335–2354.

Hansen, Lars P., and Ravi Jagannathan, 1991, "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy* 99, 225–262.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, "Digesting Anomalies: An Investment Approach," *Review of Financial Studies* 28, 650–705.

Kelly, Bryan, Seth Pruitt, and Yinan Su, 2019, "Characteristics are Covariances: A Unified Model of Risk and Return," *Journal of Financial Economics* 134, 501–524.

Kim, Soohun, Robert A Korajczyk, and Andreas Neuhierl, 2021, “Arbitrage Portfolios,” *Review of Financial Studies* 34, 2813–2856.

Kozak, Serhiy, 2019, “Kernel Trick for the Cross Section,” Working paper, University of Maryland.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2018, “Interpreting Factor Models,” *Journal of Finance* 73, 1183–1223.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2020, “Shrinking the Cross-Section,” *Journal of Financial Economics* 135, 271–292.

Lu, Cuicui, and Peter Schmidt, 2012, “Conditions for the Numerical Equality of the OLS, GLS and Amemiya–Cragg Estimators,” *Economics Letters* 116, 538–540.

Murphy, Kevin P., 2023, “Probabilistic Machine Learning: An introduction.”

Pezzo, Luca, Raja Velu, Zhaoque Zhou, and Lei Wang, 2022, “Efficient Modeling of Cross-Sectional Stock Returns via Reduced Rank (Seemingly Unrelated) Regression,” Working paper, SSRN.

Rosenberg, Barr, 1974, “Extra-Market Components of Covariance in Security Returns,” *Journal of Financial and Quantitative Analysis* 9, 263–274.

Appendix (for online publication)

A. Additional plots and figures

A.1. Empirical performance of hedged factors

In Figure A.1 we perform the exercise from Figure I using univariate factors. Interesting differences emerge from comparing results for OLS and univariate factors. First, average squared Sharpe ratio improvements decay much more slowly with the number of factors J. Second, there is higher benefit to hedging more than one round and even higher if GLS factors are used. These results suggests that univariate factors might be more contaminated with unpriced risks than OLS factors and there is more room for correcting these inefficiencies with hedging or GLS factor constructions, even for models with a large number of factors.

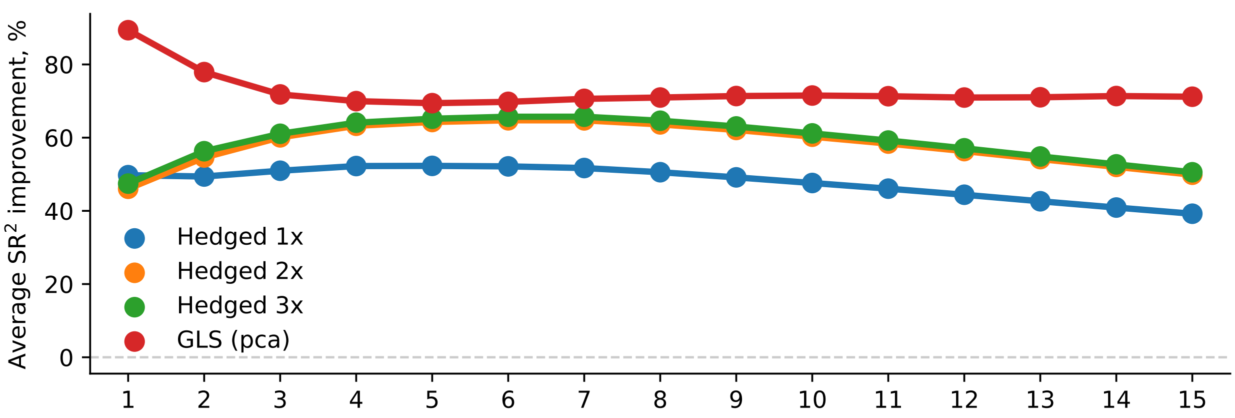

Figure A.1: **In-sample maximum squared Sharpe ratio improvement due to iterative hedging of univariate factors.** The plot shows improvement, in %, of annualized average in-sample maximum squared Sharpe ratio from hedged univariate factors relative to unhedged factors, for all models with a constant and 1–15 additional factors. We hedge the factors up to three times. We also report performance of the GLS factors which use the sample conditional covariance matrix of individual stock returns estimated using the rolling PCA procedure outlined in Section V.B. For each number of factors on the x-axis, results are averaged across 10,000 models with this number of factors randomly drawn from the set of all factors.

Figure A.2 studies orthonormalized factors. The results lie in between of OLS and univariate factors: the speed of the squared Sharpe ratio improvement decay with the number of factors J is greater than in the univariate case but lower than with OLS factors; hedging more than once provides greater benefit than for OLS factors but less than with univariate factors.

Figures A.3 and A.4 show the same results evaluated out-of-sample using the sample split approach discussed previously. Out-of-sample results exhibit similar patterns as the in-sample results, but the magnitude of effects is diminished.

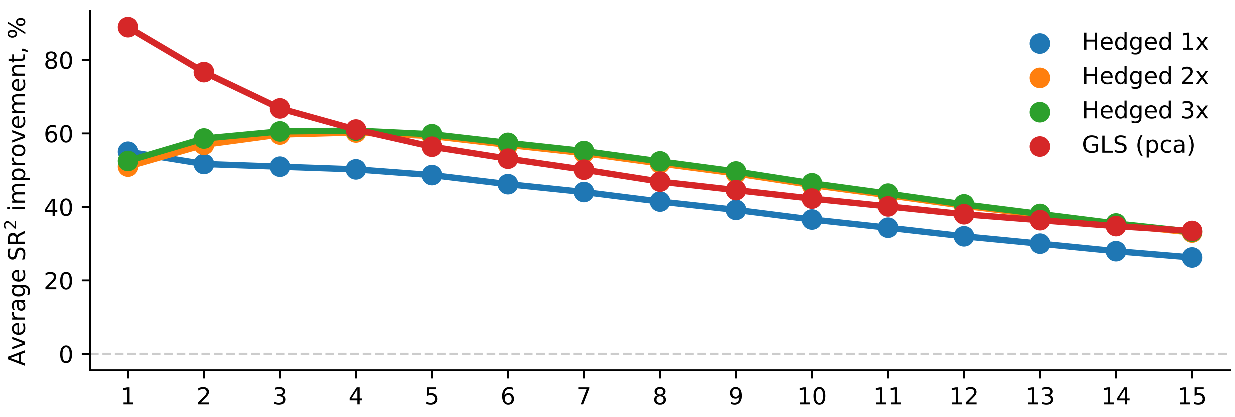

FIGURE A.2: **In-sample maximum squared Sharpe ratio improvement due to iterative hedging of orthonormalized factors.** The plot shows improvement, in %, of annualized average in-sample maximum squared Sharpe ratio from hedged orthonormalized factors relative to unhedged factors, for all models with a constant and 1–15 additional factors. We hedge the factors up to three times. We also report performance of the GLS factors which use the sample conditional covariance matrix of individual stock returns estimated using the rolling PCA procedure outlined in Section V.B. For each number of factors on the x-axis, results are averaged across 10,000 models with this number of factors randomly drawn from the set of all available factors.

Table A.I reports the level of in-sample maximum squared Sharpe ratios for hedged and GLS factors. We show results for univariate, orthonormal, and OLS factors. Table A.II shows the same results evaluated out-of-sample using the sample split approach discussed previously. Out-of-sample results exhibit similar patterns as the in-sample results, but the magnitude of effects is diminished. For OLS factors hedging provides small improvement for models with a small number of factors and no improvement for models with 12 or more factors. GLS factors (last row) work somewhat better than hedged factors and still yields some efficiency improvements even for models with 15 factors. For univariate and orthonormal factors, the improvements in squared Sharpe ratios decay slower with the number of factors. We can still see benefits for OOS performance even for models with 15 factors.

Table A.III reports maximum out-of-sample annualized squared Sharpe ratios of all two-factor models which use OLS factors (first column), OLS hedged factors for $n = 1..3$ rounds, as well as approximate GLS factors (the last column). All models include two characteristics in $\boldsymbol{X}_t$: a constant, and one of the characteristics listed in the rows.

We consider DMRS hedging as an alternative to our hedging approach. Tables A.IV and A.V show in-sample and out-of-sample results, respectively. Overall, DMRS hedging appears to be less reliable and performs significantly worse than our hedging approach.

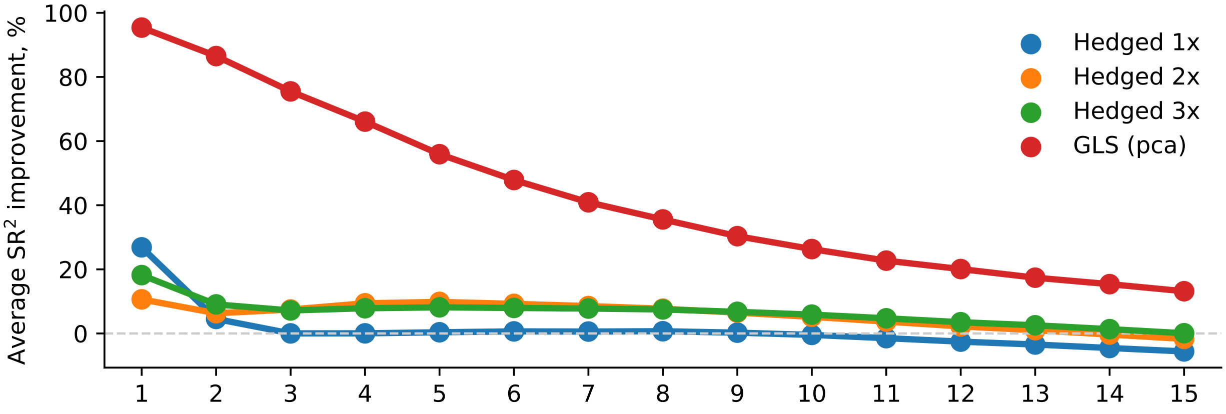

FIGURE A.3: **Out-of-sample maximum squared Sharpe ratio improvement due to iterative hedging of OLS factors.** The plot shows improvement, in %, of annualized average out-of-sample maximum squared Sharpe ratio from hedged OLS factors relative to unhedged factors, for all models with a constant and 1–15 additional factors. We hedge the factors up to three times. We also report performance of the GLS factors which use the sample conditional covariance matrix of individual stock returns estimated using the rolling PCA procedure outlined in Section V.B. For each number of factors on the x-axis, results are averaged across 10,000 models with this number of factors randomly drawn from the set of all factors.

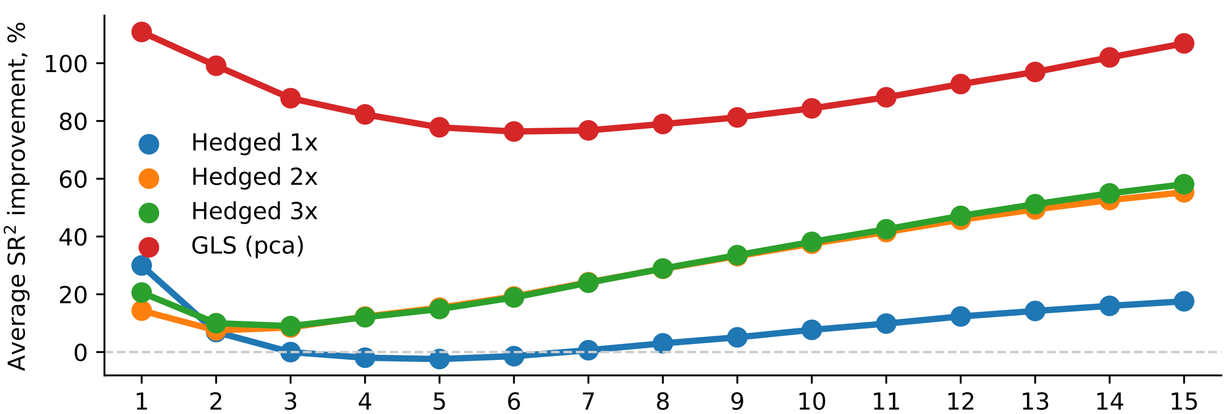

FIGURE A.4: **Out-of-sample maximum squared Sharpe ratio improvement due to iterative hedging of Univariate factors.** The plot shows improvement, in %, of annualized average out-of-sample maximum squared Sharpe ratio from hedged univariate factors relative to unhedged factors, for all models with a constant and 1–15 additional factors. We hedge the factors up to three times. We also report performance of the GLS factors which use the sample conditional covariance matrix of individual stock returns estimated using the rolling PCA procedure outlined in Section V.B. For each number of factors on the x-axis, results are averaged across 10,000 models with this number of factors randomly drawn from the set of all factors.

TABLE A.I: In-sample maximum squared Sharpe ratios of hedged factors.

We construct hedged factors and iterate by hedging up to three times. The table shows average in-sample annualized maximum squared Sharpe ratios from hedged univariate, orthonormal, or OLS factors (panels) relative to unhedged factors (first row in each panel), in %, for all models with a constant and 1–15 additional factors. We also report performance of the GLS factors which use the sample conditional covariance matrix of individual stock returns estimated using the rolling PCA procedure outlined in Section V.B (last row). For each number of factors reported in the columns, results are averaged across 10,000 models with this number of factors randomly drawn from the set of all available factors.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
								Univariate							
Unhedged	0.8	1.1	1.5	1.9	2.3	2.8	3.2	3.7	4.2	4.7	5.2	5.7	6.2	6.6	7.1
Hedged 1x	1.2	1.6	2.2	2.9	3.5	4.2	4.9	5.6	6.3	6.9	7.6	8.2	8.8	9.3	9.9
Hedged 2x	1.1	1.6	2.3	3.1	3.8	4.6	5.3	6.1	6.8	7.5	8.2	8.9	9.5	10.1	10.6
Hedged 3x	1.1	1.6	2.3	3.1	3.8	4.6	5.4	6.1	6.9	7.6	8.3	8.9	9.5	10.1	10.7
								Orthonormal							
Unhedged	0.8	1.1	1.5	2.0	2.5	3.1	3.7	4.3	5.0	5.7	6.3	7.0	7.7	8.4	9.1
Hedged 1x	1.2	1.6	2.3	3.0	3.7	4.5	5.3	6.1	6.9	7.7	8.5	9.3	10.0	10.8	11.5
Hedged 2x	1.2	1.7	2.4	3.2	4.0	4.8	5.7	6.6	7.4	8.3	9.1	9.9	10.6	11.4	12.1
Hedged 3x	1.2	1.7	2.4	3.2	4.0	4.9	5.7	6.6	7.5	8.3	9.1	9.9	10.7	11.4	12.1
								OLS							
Unhedged	0.8	1.0	1.5	2.0	2.6	3.3	4.0	4.7	5.5	6.3	7.1	7.9	8.8	9.6	10.4
Hedged 1x	1.2	1.6	2.3	3.0	3.8	4.6	5.4	6.3	7.1	8.0	8.8	9.6	10.4	11.2	12.0
Hedged 2x	1.2	1.7	2.4	3.2	4.0	4.9	5.8	6.6	7.5	8.3	9.2	10.0	10.8	11.5	12.3
Hedged 3x	1.2	1.7	2.4	3.2	4.0	4.9	5.8	6.7	7.5	8.4	9.2	10.0	10.8	11.6	12.3
GLS (pca)	1.5	1.9	2.5	3.2	3.9	4.7	5.5	6.4	7.2	8.1	8.9	9.7	10.5	11.4	12.1

TABLE A.II: Out-of-sample maximum squared Sharpe ratios of hedged factors.

We construct hedged factors and iterate by hedging up to three times. The table shows average out-of-sample annualized maximum squared Sharpe ratios from hedged Univariate, Orthonormal, or OLS factors (panels) relative to unhedged factors (first row in each panel), in %, for all models with a constant and 1–15 additional factors. We also report performance of the GLS factors which use the sample conditional covariance matrix of individual stock returns estimated using the rolling PCA procedure outlined in Section V.B (last row). For each number of factors reported in the columns, results are averaged across 10,000 models with this number of factors randomly drawn from the set of all available factors.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
								Univariate							
Unhedged	0.4	0.5	0.5	0.6	0.7	0.8	0.8	0.9	0.9	1.0	1.0	1.0	1.1	1.1	1.1
Hedged 1x	0.5	0.5	0.5	0.6	0.7	0.8	0.8	0.9	1.0	1.0	1.1	1.2	1.2	1.3	1.3
Hedged 2x	0.5	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
Hedged 3x	0.5	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
								Orthonormal							
Unhedged	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
Hedged 1x	0.5	0.5	0.6	0.7	0.7	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
Hedged 2x	0.5	0.5	0.6	0.7	0.9	1.0	1.1	1.2	1.4	1.5	1.6	1.7	1.8	1.9	2.0
Hedged 3x	0.5	0.5	0.6	0.7	0.8	1.0	1.1	1.2	1.4	1.5	1.6	1.7	1.8	2.0	2.1
								OLS							
Unhedged	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.3	1.4	1.5	1.7	1.8	1.9	2.1
Hedged 1x	0.5	0.5	0.6	0.7	0.8	0.9	1.0	1.2	1.3	1.4	1.5	1.6	1.7	1.8	2.0
Hedged 2x	0.5	0.5	0.6	0.8	0.9	1.0	1.1	1.2	1.4	1.5	1.6	1.7	1.8	1.9	2.0
Hedged 3x	0.5	0.5	0.6	0.7	0.9	1.0	1.1	1.2	1.4	1.5	1.6	1.7	1.8	2.0	2.1
GLS (pca)	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.4

TABLE A.III: Out-of-sample maximum squared Sharpe ratios of two-factor OLS models. We report maximum out-of-sample annualized squared Sharpe ratios of all models which use OLS factors (first column), OLS hedged factors for $n = 1..3$ rounds, as well as approximate GLS factors (the last column). All models include two characteristics in $\boldsymbol{X}_t$: a constant, and one of the characteristics listed in the rows. GLS factors use a non-parametric covariance matrix estimated via PCA applied to 3-year rolling windows of daily stocks returns. The row labeled "ER" uses fitted values from a panel regression of returns on all characteristics as a standalone characteristic. The last row averages the numbers across all models. Out-of-sample results are based on a split sample estimation before/after 2005.

	OLS	Hedged n times			GLS
		1	2	3	
Size	0.26	0.65	0.31	0.42	0.66
Value (A)	0.20	0.08	0.11	0.12	0.43
Gross Profitability	0.64	1.03	0.97	1.10	1.12
F-score	0.50	1.05	1.14	1.38	1.51
Debt Issuance	0.40	0.62	0.57	0.49	0.85
Share Repurchases	0.44	0.55	0.45	0.61	0.74
Net Issuance (A)	0.70	0.85	0.81	0.94	0.91
Asset Growth	0.26	0.23	0.18	0.19	0.55
Asset Turnover	0.67	0.77	0.57	0.50	0.72
Gross Margins	0.40	0.73	0.43	0.56	0.77
Earnings/Price	0.40	0.64	0.52	0.66	0.82
Investment/Capital	0.31	0.27	0.20	0.30	0.75
Investment Growth	0.30	0.18	0.22	0.21	0.57
Sales Growth	0.41	0.22	0.24	0.24	0.62
Leverage	0.19	0.07	0.09	0.10	0.41
Return on Assets (A)	0.45	0.95	0.76	0.84	1.00
Return on Book Equity (A)	0.42	0.98	0.64	0.79	0.91
Sales/Price	0.36	0.51	0.37	0.40	0.64
Momentum (6m)	0.27	0.39	0.42	0.43	0.73
Industry Momentum	0.85	0.72	0.51	0.44	1.22
Momentum (12m)	0.41	0.39	0.47	0.44	0.70
Momentum-Reversals	0.24	0.43	0.43	0.37	0.54
Value (M)	0.21	0.25	0.37	0.36	0.66
Net Issuance (M)	0.77	1.02	0.80	0.81	2.01
Short-Term Reversals	0.12	0.18	0.19	0.21	0.40
Idiosyncratic Volatility	0.56	0.37	0.18	0.23	1.03
Beta Arbitrage	0.90	0.59	0.62	0.63	1.03
Industry Rel. Reversals	0.08	0.19	0.22	0.22	0.40
Price	0.34	0.47	0.54	0.51	0.72
Firm's age	0.38	0.77	0.53	0.56	0.74
Share Volume	0.66	0.54	0.37	0.36	1.14
Exchange Switch	1.17	2.08	2.00	2.03	1.70
IPO	0.30	0.59	0.33	0.35	0.62
ER	1.31	1.34	1.58	1.61	2.56
Average	0.47	0.61	0.53	0.57	0.89

TABLE A.IV: In-sample maximum squared Sharpe ratios for DMRS-hedged factors.

We construct hedged factors and iterate by hedging using the DMRS procedure up to three times. The table shows annualized average in-sample maximum squared Sharpe ratios from hedged univariate, orthonormal, or OLS factors (panels) relative to unhedged factors (first row in each panel), in %, for all models with a constant and 1–15 additional factors. We also report performance of the GLS factors which use the sample conditional covariance matrix of individual stock returns estimated using the rolling PCA procedure outlined in Section V.B (last row). For each number of factors reported in the columns, results are averaged across 10,000 models with this number of factors randomly drawn from the set of all available factors.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
								Univariate							
Unhedged	0.8	1.1	1.4	1.9	2.3	2.8	3.2	3.7	4.2	4.7	5.2	5.6	6.1	6.6	7.1
Hedged 1x	0.9	1.3	1.7	2.2	2.7	3.1	3.6	4.1	4.6	5.1	5.6	6.0	6.5	6.9	7.4
Hedged 2x	1.0	1.3	1.8	2.3	2.7	3.2	3.7	4.2	4.6	5.1	5.6	6.0	6.5	6.9	7.4
Hedged 3x	1.0	1.4	1.8	2.3	2.8	3.2	3.7	4.2	4.6	5.1	5.6	6.0	6.5	6.9	7.3
								Orthonormal							
Unhedged	0.8	1.1	1.5	2.0	2.5	3.1	3.7	4.4	5.0	5.7	6.4	7.0	7.8	8.4	9.1
Hedged 1x	0.9	1.3	1.8	2.3	2.8	3.5	4.1	4.8	5.4	6.1	6.7	7.3	8.0	8.6	9.2
Hedged 2x	1.0	1.4	1.9	2.4	2.9	3.5	4.1	4.8	5.4	6.1	6.7	7.3	8.0	8.6	9.2
Hedged 3x	1.0	1.4	1.9	2.4	2.9	3.5	4.1	4.8	5.4	6.0	6.7	7.3	7.9	8.5	9.1
								OLS							
Unhedged	0.8	1.0	1.5	2.0	2.6	3.3	4.0	4.7	5.5	6.3	7.1	8.0	8.8	9.6	10.5
Hedged 1x	0.9	1.3	1.8	2.4	3.0	3.6	4.3	5.1	5.8	6.6	7.4	8.1	8.9	9.7	10.5
Hedged 2x	1.0	1.3	1.9	2.4	3.0	3.7	4.4	5.1	5.9	6.6	7.4	8.1	8.9	9.7	10.4
Hedged 3x	1.0	1.4	1.9	2.4	3.1	3.7	4.4	5.1	5.8	6.6	7.3	8.1	8.9	9.6	10.4
GLS (pca)	1.5	1.9	2.5	3.2	4.0	4.8	5.6	6.5	7.3	8.1	9.0	9.8	10.6	11.4	12.2

TABLE A.V: Out-of-sample maximum squared Sharpe ratios for DMRS-hedged factors.

We construct hedged factors and iterate by hedging using the DMRS procedure up to three times. The table shows annualized average out-of-sample maximum squared Sharpe ratios from hedged univariate, orthonormal, or OLS factors (panels) relative to unhedged factors (first row in each panel), in %, for all models with a constant and 1–15 additional factors. We also report performance of the GLS factors which use the sample conditional covariance matrix of individual stock returns estimated using the rolling PCA procedure outlined in Section V.B (last row). For each number of factors reported in the columns, results are averaged across 10,000 models with this number of factors randomly drawn from the set of all available factors.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
								Univariate							
Unhedged	0.4	0.5	0.6	0.6	0.7	0.8	0.8	0.9	0.9	1.0	1.0	1.0	1.1	1.1	1.1
Hedged 1x	0.4	0.5	0.6	0.7	0.7	0.8	0.8	0.9	0.9	0.9	1.0	1.0	1.1	1.1	1.2
Hedged 2x	0.4	0.5	0.6	0.7	0.7	0.8	0.8	0.9	0.9	0.9	1.0	1.0	1.1	1.1	1.1
Hedged 3x	0.4	0.5	0.6	0.7	0.7	0.8	0.8	0.8	0.9	0.9	1.0	1.0	1.0	1.1	1.1
								Orthonormal							
Unhedged	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
Hedged 1x	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7
Hedged 2x	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7
Hedged 3x	0.4	0.5	0.6	0.7	0.8	0.9	0.9	1.0	1.1	1.2	1.2	1.3	1.4	1.5	1.6
								OLS							
Unhedged	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.3	1.4	1.5	1.7	1.8	1.9	2.1
Hedged 1x	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.6	1.7	1.8	1.9
Hedged 2x	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.7	1.8	1.9
Hedged 3x	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.9
GLS (pca)	0.8	0.9	1.0	1.1	1.2	1.4	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.4

TABLE A.VI: Maximum squared Sharpe ratios of hedged factors (DMRS hedging).
We report in-sample (top panel) and out-of-sample (bottom panel) annualized maximum squared Sharpe ratio of the MVE portfolio constructed from 34 unhedged (first column) or DMRS-hedged (up to five times) factors. Rows correspond to three types of factors we discuss in Section V.A.

	Unhedged	Hedged n times				
		1	2	3	4	5
	In-sample					
Univariate	13.8	13.6	13.4	13.2	13.2	13.2
Orthonormal	18.0	17.8	17.5	17.5	17.5	17.5
OLS	21.3	21.0	20.8	20.8	20.9	20.9
	Out-of-sample					
Univariate	1.3	1.5	1.4	1.3	1.3	1.3
Orthonormal	3.4	3.4	3.1	3.1	3.1	3.1
OLS	4.0	4.3	4.2	4.2	4.2	4.2

In Table A.VI we consider the effect of hedging on the models with the full set of 34 factors, both in sample (top panel) and out of sample (bottom panel). The Table reports annualized maximum squared Sharpe ratio of the mean-variance optimal portfolio constructed from 34 unhedged (first column) or DMRS-hedged (up to five times) factors. Rows correspond to three types of factors we discuss in Section V.A.

A.2. Dimensionality reduction

In Table A.VII we investigate whether the performance of latent factors can be improved by using our hedging procedure. Because latent factors are designed to explain as much variation in realized returns as possible, for a given number of factors, we would expect the violations of the conditions of Proposition 2 to be quantitatively less important for latent factor models with a sufficient number of factors. The table presents results for the four type of latent factors models we considered previously (shown in their respective panels). For each type, the first row in a panel shows maximum squared Sharpe ratios from unhedged models with 1..12 factors (rows), the three subsequent rows show results from hedging the latent-factor model's implied weights using our hedging procedure up to three times. * and ** indicate significance of the squared Sharpe ratio difference between the given model and the unhedged benchmark (first row) at the 5% and 1% levels, respectively.

We find that hedging can improve the performance of latent factor models, especially the ones that are not OLS transformed, such as the SCS model. The improvements for PPCA are also sizeable and statistically significant in this dataset. For OLS transformed models such as IPCA, improvements are more muted and largely insignificant for models with a sufficient number of factors. Table A.VIII shows that these findings translate to out-of-sample data.

As expected, the GLS version of IPCA factors is more efficient than standard IPCA and

TABLE A.VII: In-sample hedging of latent factor models.

The table shows the effect of hedging on four latent factor models (panels). The first row in each panel shows annualized in-sample maximum squared Sharpe ratios of the model with $n = 1..12$ factors (columns). The three subsequent rows hedge this model iteratively. * indicate $p < 0.05$ of a one-sided test of the squared Sharpe ratio difference of the given model relative to the unhedged benchmark (first row). ** indicate $p < 0.01$.

	1	2	3	4	5	6	7	8	9	10	11	12
						SCS						
Unhedged	0.2	0.6	0.9	1.2	3.1	3.1	3.1	4.4	4.7	4.7	7.9	8.1
Hedged 1x	0.1	0.6	1.3	1.7	7.4**	7.5**	7.6**	9.1**	9.2**	9.3**	10.9**	10.9**
Hedged 2x	0.1	0.6	1.2	1.8*	7.6**	7.8**	7.9**	9.6**	9.7**	9.8**	11.4**	11.5**
Hedged 3x	0.1	0.6	1.3*	1.8*	7.6**	7.8**	7.8**	9.6**	9.7**	9.8**	11.5**	11.6**
						IPCA						
Unhedged	0.3	1.3	4.1	4.5	7.0	7.7	11.9	12.6	13.7	14.4	14.8	15.3
Hedged 1x	0.4	1.4	4.6	5.1	6.8	7.7	12.7	13.4	15.1	15.8	16.0	16.2
Hedged 2x	0.3	1.4	5.0*	5.6*	7.4	8.2	13.3	13.7	15.1	15.9	16.0	16.3
Hedged 3x	0.4	1.5	4.9	5.6*	7.4	8.2	13.4	13.7	15.1	15.8	15.9	16.2
						PPCA						
Unhedged	0.3	0.3	0.7	2.5	8.3	8.3	8.7	12.0	12.0	13.2	13.2	13.3
Hedged 1x	0.3	0.3	1.1*	3.9**	12.4**	12.7**	13.0**	14.5**	15.0**	15.8**	16.0**	16.0**
Hedged 2x	0.3	0.3	1.1	4.1**	12.6**	12.9**	13.3**	14.8**	15.0**	15.7*	16.0**	16.1**
Hedged 3x	0.4	0.3	1.1*	4.1**	12.6**	12.8**	13.2**	14.8**	15.0**	15.6*	16.0**	16.0**
						IPCA (GLS)						
Unhedged	0.6	1.3	11.1	10.9	12.0	12.9	16.4	16.8	16.7	16.5	16.3	16.4
Hedged 1x	0.6	1.3	11.1	10.8	12.0	12.8	16.5	16.9	16.8	16.5	16.3	16.3
Hedged 2x	0.6	1.3	11.0	10.7	11.9	12.8	16.5	17.0	16.9	16.6	16.4	16.4
Hedged 3x	0.6	1.3	11.0	10.7	11.9	12.8	16.5	16.9	16.8	16.5	16.3	16.4

TABLE A.VIII: Out-of-sample hedging of latent factor models.

The table shows the effect of hedging on four latent factor models (panels). The first row in each panel shows annualized out-of-sample maximum squared Sharpe ratios of the model with $n = 1..12$ factors (columns). The three subsequent rows hedge this model iteratively. Out-of-sample results are based on a split sample estimation before/after 2005.

	1	2	3	4	5	6	7	8	9	10	11	12
						SCS						
Unhedged	0.1	0.2	0.4	0.5	0.4	0.3	0.3	0.6	0.8	0.8	1.6	1.5
Hedged 1x	0.0	0.1	0.4	0.4	0.9	0.7	0.7	1.3	1.4	1.4	2.7	2.7
Hedged 2x	0.0	0.1	0.5	0.5	1.1	1.0	0.9	1.6	1.6	1.6	2.9	2.9
Hedged 3x	0.0	0.2	0.6	0.5	1.0	0.9	0.9	1.6	1.6	1.6	3.0	3.0
						IPCA						
Unhedged	0.3	0.1	0.7	0.8	1.0	1.1	2.1	2.2	2.6	3.1	3.5	3.8
Hedged 1x	0.2	0.1	0.3	0.4	0.5	0.8	2.1	2.5	3.6	4.3	4.3	4.3
Hedged 2x	0.1	0.1	0.3	0.5	0.8	1.2	2.8	2.9	4.1	4.5	4.2	4.3
Hedged 3x	0.3	0.1	0.3	0.5	0.9	1.1	2.7	2.8	4.0	4.4	4.2	4.2
						PPCA						
Unhedged	0.2	0.2	0.4	1.0	1.6	1.3	1.2	3.0	2.4	3.2	3.1	3.1
Hedged 1x	0.1	0.1	0.6	1.0	2.9	2.3	2.3	3.4	3.6	4.0	4.0	4.0
Hedged 2x	0.0	0.1	0.6	0.9	3.0	2.4	2.6	3.6	3.6	4.0	4.1	4.1
Hedged 3x	0.2	0.1	0.6	0.9	2.9	2.3	2.5	3.6	3.6	4.0	4.1	4.1
						IPCA (GLS)						
Unhedged	0.4	0.2	2.7	2.2	2.8	2.8	4.7	4.8	3.8	3.8	3.8	3.7
Hedged 1x	0.4	0.2	2.7	2.2	2.9	2.8	4.8	4.9	4.0	3.9	3.8	3.7
Hedged 2x	0.4	0.2	2.7	2.1	2.8	2.7	4.7	4.8	3.9	3.9	3.7	3.6
Hedged 3x	0.4	0.2	2.7	2.1	2.8	2.8	4.7	4.8	3.9	3.9	3.7	3.6

does not benefit from hedging.

B. An alternative dataset

In this section we use a different dataset with a much wider selection of characteristics. The data is based on Wharton Research Data Services "Backtester Plus" dataset from the Factors by WRDS suite. It contains 134 signals based on CRSP Stocks, Compustat, IBES, OptionMetrics, Thomson Reuters, and WRDS SEC Analytics databases. The entire list of factors is available on the WRDS website.

We rank-transform and standardize stock characteristics from this dataset and merge them with daily stock returns from CRSP. The sample is from January 1975 to December 2020. We apply several filters to preserve characteristics with maximum data availability. In particular, we remove binary characteristics and any characteristics for which more than 25% of the observations in the panel of firms are missing. We remove any time periods in the early part of the sample for which less than 500 firms are available. We also remove firms whose past market caps do not exceed 0.01% of the aggregate stock market capitalization (e.g., firms with market capitalizations less than $4 billion on a $40 trillion aggregate stock market valuation).

Importantly, instead of filling in any missing characteristics with their cross-sectional means as we did in our main exercise, we impute characteristic values using an advanced imputation method based on Huang and Kozak (2023). [20]

The resulting dataset contains 107 months of monthly characteristics and daily returns on 4,825 stocks.

Figures A.5 and A.6 below report results for this dataset.

20. They develop a Bayesian tensor model to impute missing or infrequently observed financial data on firm characteristics. One of the advantages of their setup is that they model and use the time-series and cross-sectional dependencies of firm characteristics in a unified and flexible way, which significantly improves imputation accuracy and allows for statistical inference via *multiple imputation* by averaging over random samples of missing characteristics drawn from the joint probability distribution they estimate.

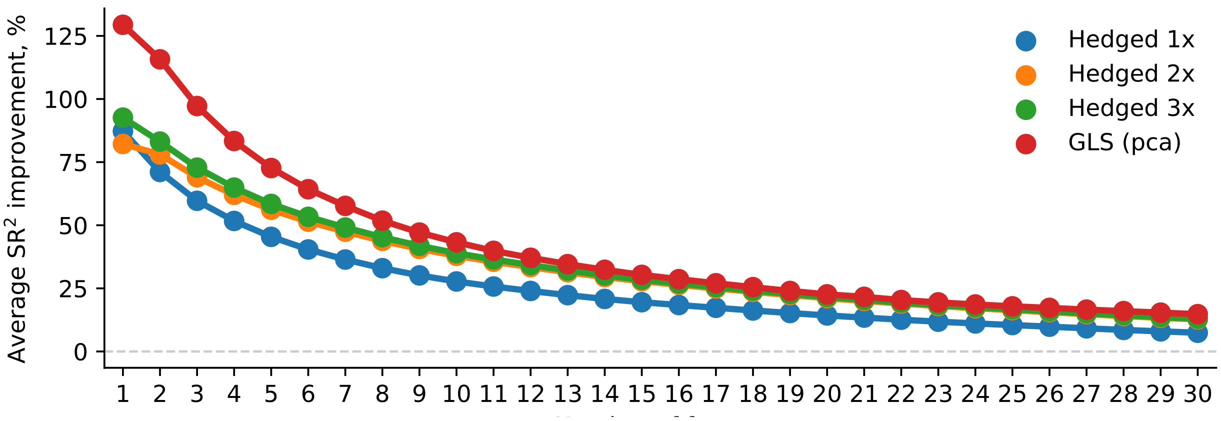

FIGURE A.5: **In-sample maximum squared Sharpe ratio improvement due to iterative hedging of OLS factors.** The plot shows improvement, in %, of annualized average in-sample maximum squared Sharpe ratio from hedged OLS factors relative to unhedged factors, for all models with a constant and 1–15 additional factors. We hedge the factors up to three times. We also report performance of the GLS factors which use the sample conditional covariance matrix of individual stock returns estimated using the rolling PCA procedure outlined in Section V.B. For each number of factors on the x-axis, results are averaged across 10,000 models with this number of factors randomly drawn from the set of all factors.

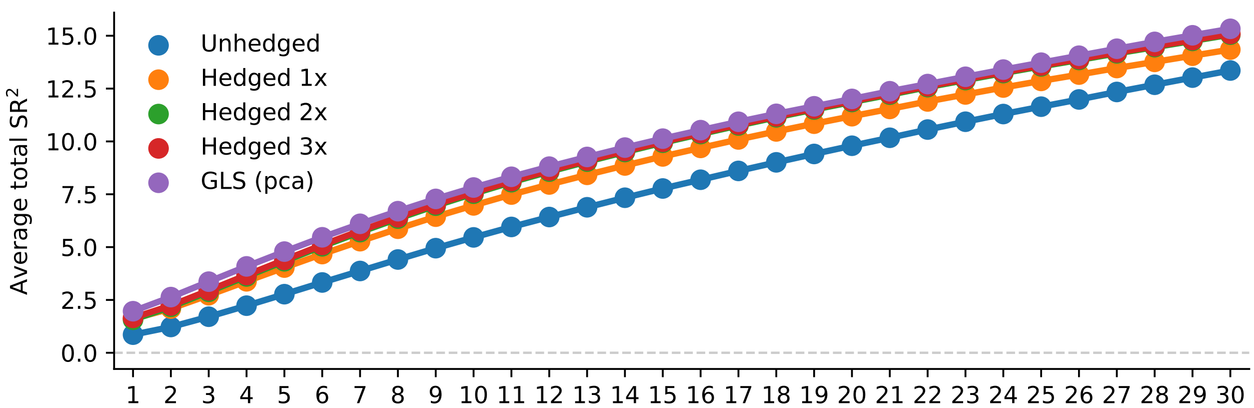

FIGURE A.6: **In-sample maximum squared Sharpe ratio of hedged OLS factors.** The plot shows annualized average in-sample maximum squared Sharpe ratios of unhedged and hedged OLS factors, as well as GLS factors. We hedge the factors up to three times. The latter use the sample conditional covariance matrix of individual stock returns estimated using the rolling PCA procedure outlined in Section V.B. For each number of factors on the x-axis, results are averaged across 10,000 models with this number of factors randomly drawn from the set of all factors.