Empirical Asset Pricing with Functional Factors*

Philip Nadler† Alessio Sancetta‡

February 4, 2022

Abstract

We propose a methodology to use functional factors in empirical asset pricing models. We establish conditions under which it is possible to recover linear beta pricing. The proposed estimation approach allows us to use high dimensional functional curves, such as the term structure of interest rates or the implied volatility smile, as factors. This framework enables the estimation of functional factor loadings as well as risk premium parameters of factor models. We derive estimation algorithms and establish the asymptotic consistency and normality of the parameter estimates. In an empirical application, we show that the implied variance smile of the S&P500 is a potential pricing factor for momentum sorted portfolios. In particular, a positive risk premium is earned by the convexity of the implied variance curve.

Key Words: bootstrap, functional data analysis, functional risk premium, implied volatility curve.

JEL Codes: G12, C13

*We are greatly indebted to the Editor Fabio Trojani, the Associate Editor and the Referee for comments that have led to substantial improvements both in content and presentation. We also thank Rodrigo Dupleich, Andreas Kaeck, Alex Kurov, Steve Satchell, and the seminar participants at Nottingham Business School as well as the University of Sussex Young Finance Scholar Conference for helpful comments on earlier drafts of the paper.

†Department of Computing, Imperial College London, London SW7 2AZ, UK; email: p.nadler@imperial.ac.uk.

‡Department of Economics, Royal Holloway University of London, Egham TW20 0EX, UK; email: asancetta@gmail.com.

1 Introduction

This paper studies empirical asset pricing models in which risky assets load on functional factors. Typical examples are the term structure of interest rates and the option volatility smile. Formally, a functional factor takes values in a normed space, possibly infinite dimensional. Here we focus on Hilbert valued functional factors. An early example of the regularity of these factors and the associated loadings is Cochrane and Piazzesi (2005). Using a vector of term structure yields at different tenors, they find that the associated loadings have a remarkably regular structure. Many economic variables of interest exhibit a functional form. A functional framework has the benefits of enhancing statistical efficiency for parameter estimation by viewing estimation in terms of one functional parameter instead of multiple scalar parameters. For the sake of definiteness, let us recall the definition of pricing factor when the factor is a vector. To this end, let R_t be a $K \times 1$ vector of returns on risky assets at time t. The $L \times 1$ vector F_t is a pricing factor for the K risky assets if there is a constant α and an $L \times 1$ vector λ such that

$$\mathbb{E}R_t = 1_K \alpha + \beta\lambda$$

where $\beta = Cov\left(R_t, F_t\right) Var\left(F_t\right)^{-1}$ and 1_K is the $K \times 1$ vector with entries equal to one. Throughout, the symbol $'$ stands for transpose, $Var\left(F_t\right)$ is the variance of the vector F_t, an $L \times L$ dimensional matrix, and $Cov\left(R_t, F_t\right)$ is the covariance of the arguments, a $K \times L$ matrix (Munk, 2013, Definition 10.1). For an economy with a risk free interest rate and one pricing factor F_t equal to the market return, we recover the Capital Asset Pricing Model (CAPM). In this case, α is the risk free interest rate and β is the vector of market betas. We extend this definition to functional factors and give conditions for its validity. In particular, let $F_t := \{F_t(s) : s \in [0,1]\}$ be an $L \times 1$ vector valued continuous function. We give conditions for F_t to be a functional pricing factor for the K risky assets in terms of the following pricing equation

$$\mathbb{E}R_t = 1_K \alpha + \int_0^1 \beta(s)\,\lambda(s)\,ds. \tag{1}$$

Here, α is a constant, β is a $K \times L$ matrix valued function of loadings and λ is an L-dimensional column vector of functional risk premia. The exact meaning of (1) is given in Definition 1, in Section 2. When $F_t(s)$ is constant for each s we are back to the usual framework of pricing factors. For $L = 1$, if F_t is a term structure pricing factor based on yields at different maturities, then we can regard F_t as a function of the tenor standardised to be in $[0, 1]$. In our empirical application, we consider the case $L = 2$, where F_t is the return on the market portfolio together with the unpredictable part of the 3-month options implied volatility smile on the S&P500.

The goal of the paper is to extend empirical asset pricing to functional factors. We show the restrictions that need to be satisfied by α and β in (1) in order to guarantee that F_t is a pricing factor. Assuming that all quantities are Hilbert space valued, we provide representations that can be used to decompose risk into factor scores. In practice, to recover the factor scores that are priced, we use standard principal components. This paper shows that the econometric procedure is valid under weak conditions.

To our knowledge, this is the first study concerned with this important aspect of asset pricing. The availability of large datasets has led to considerable proliferation of empirical asset pricing models via the so called factor zoo (Cochrane, 2011). However, with the exception of a few studies (e.g. Cochrane and Piazzesi, 2005, Ait-Sahalia et al., 2018, Della Corte et al., 2021, and references therein), the information coming from functional factors such as term structures has not been exploited.

1.1 Functional Factors and Principal Components

Here we use the standard term structure of interest rates to showcase the importance of functional factors. Consider the yield on a zero coupon government bond with maturity (tenor) j months from now. For example, we may observe yields at 3 months up to 30 years. In this case, suppose that we use the linear transformation $s(j) = (j - 3)/360$ where j is the tenor. In our context the yield on a 12 months zero coupon bond is written as $F_t(s(12))$ $(j = 12)$. Assuming that we observe yields on zero coupon bonds for every month, $\{F_t(s(j)) : j = 3, 4, 5, ..., 360\}$ is a vector of dimension 357. It is customary to use principal components and only extract the first three factor scores to be used as pricing factors. The first three factor scores are usually interpretable as level, slope and curvature. The possible problem with this procedure is that a consistent estimator of the covariance matrix is required in order to compute principal components in a consistent way (Johnstone and Lu, 2009). With no further restrictions, consistency of a 357×357 dimensional covariance matrix requires sample sizes that are prohibitively large. However, in a functional data framework the estimation of large covariance matrices becomes a standard procedure. Essentially, this is the case if $\int_0^1 \mathbb{E}F_t^2(s)\,ds < \infty$, switching from summations to integrals when we assume that we can observe yields at a continuous number of expiry dates. From a mathematical point of view the switch to continuous $s \in [0,1]$ removes the dependency on the dimension. Hence, the functional factor framework allows us to justify commonly used approaches in empirical asset pricing when the data are high dimensional, but with some structure. This is not just a statistical problem as it has theoretical implications for pricing and the validity of (1). In this respect, we provide the theoretical details for pricing in this general framework (Section 2).

Within the functional pricing framework, interest does not only exclusively lie in estimating factor scores. We may want to know if there is a specific portion of the yield curve with higher loading relative to another. Similarly, we may want to know what portion of the yield curve produces a higher risk premium. These questions can be answered within the framework of this paper considering functional loadings and risk premia.

In summary the functional data framework allows us to cover problems where factors can be represented as curves (possibly partially observed) in a unified a consistent way irrespective of the number of points at which the curve is observed.

1.2 Outline of the Paper

The rest of the paper is structured as follows. Section 2 introduces the definition of pricing functional factor and the conditions that need to be satisfied in this context (Definition 1 and Lemmas 1 and 2). Section 2.6 introduces the estimation algorithms. Consistency and asymptotic normality of the estimators can be found in Section 2.7. Section 3 discusses the application to the implied variance curve. In particular, we show that the curvature of implied volatility helps in explaining the risk premium earned by momentum strategies. Section 4 concludes. The proof of all the results including additional details concerning the empirical analysis can be found in the Supplementary Material to this paper (Sections A.1 and A.3). There we also include a simulation study to assess the finite sample performance of the estimators under realistic simulation designs (Section A.4).

2 Functional Pricing Factors

2.1 Scope and Limitations

Functional data has found numerous application in econometrics (inter alia, Ramsay and Ramsey, 2002, Kargin and Onatski, 2008, Müller et al., 2011, Kokoszka et al., 2015, Sancetta, 2015, 2019). The statistics literature is vast (Wang et al., 2016, for an article review and Bosq, 2000, Horváth and Kokoszka, 2012, and Ferraty and Vieu, 2016, for monograph treatments).

We briefly elaborate on some details within the context of the paper. To simplify the discussion, suppose that the factor $F_t(s)$ is real valued (i.e. $L = 1$), mean zero, and continuous in $s \in [0, 1]$. We wish to represent this in terms of a few factor scores, as we would in the case of vector valued factors, using principal components. Even if F_t is a function, hence infinite dimensional, we can find good approximations using only a few factor scores if the covariance function $C_{FF}(s, v) := \mathbb{E}F_t(s) F_t(v)$, $s, v \in [0, 1]$, is the kernel of a compact operator. This is the case if $\mathbb{E}\int_0^1 F_t^2(s)\, ds < \infty$. For simplicity we shall use the same symbol for the operator and its kernel. By the Karhunen-Loeve expansion, we have that $F_t(s) = \sum_{i=1}^{\infty} \sqrt{\rho_i}\xi_{t,i}\Phi_i(s)$ where $(\rho_i)_{i\geq 1}$ and $(\Phi_i)_{i\geq 1}$ are eigenvalues and eigenfunctions of the C_{FF}, and equality is under the uniform norm if C_{FF} is continuous. The condition that $\mathbb{E}\int_0^1 F_t^2(s)\, ds < \infty$ not only means that C_{FF} is compact, but also that $(\rho_i)_{i\geq 1}$ is summable. Then, we cannot invert C_{FF} because its eigenvalues converge to zero. However, we can construct factor scores $S_{t,i} = \int_0^1 F_t(s)\, \Phi_i(s)\, ds = \sqrt{\rho_i}\xi_{t,i}$, $i = 1, 2, ..., I$ for some finite I. As long as the I^{th} largest eigenvalue ρ_I is strictly greater than zero, the covariance matrix of the first I factor scores is invertible. If the eigenvalues decay fast, we can expect good approximations for small I.

For clarity and to put the discussion into context, we provide high level details on what type of problems we cover within this framework. The paper presents conditions for pricing using functional factors. In this respect, the context is the one of possibly large number K of risky assets but a finite number of L functional factors. While the number L of functional factor is finite, we allow it to be greater than one. This is important in order to provide a general treatment. For example, we can have standard real valued factors like the market returns and functional factors as pricing factors. In this case, we can think of $L = 2$ and the the real valued factor to be a constant function of its argument $s \in [0, 1]$. Finally, the use of functional pricing factors requires some care because of the infinite dimensionality of the pricing factors. The details will be provided in Definition 1 and Lemmas 1 and 2.

Most of the quantities used in empirical asset pricing need to be estimated. This paper provides the justification for doing so within the standard principal components approach. The econometric results of this paper include the case when the dimension K of assets is large. Mathematically, this is addressed by letting $K \to \infty$. We do this in a way that does not depend on the sample size. This is important because the number of assets to price can be large and we should not be restricted by the sample size. We do place additional restrictions on the number of priced factor scores in order to use standard estimation methods. For the moment, suppose again that $L = 1$. Then, we suppose that only a finite number I of factor scores is priced in order to provide consistency and asymptotic normality of the estimators. Given that R_t is K dimensional, it is clear that (1) is trivially satisfied for I large enough if K is a fixed finite number. This is no different from the case where we have L uncorrelated scalar factors and we let $L \to \infty$. For this reason, we confine attention to finite and bounded I. To put the restriction into perspective, in the case of the term structure of interest rate, the first three factor scores are customary used. These can be interpreted as level, slope and curvature. In our empirical application, we consider the implied volatility curve and only work with the first three factor scores. These explain 99% of variability. Hence, the scope is the one of a few functional factors pricing a large number of risky assets. Within these functional factors, only the risk from a small number of factor scores is priced. We also note that I can be small but unknown. In this case, the usual methods for selecting the number of principal components apply (e.g. Gavish and Donoho, 2014, and references therein). One of our theoretical results (Theorem 2) also suggests that we can directly test for the number of non-priced factor loadings.

Finally, we work with densely observed data. In consequence, we use the standard empirical covariance function estimator, which is consistent under the conditions we shall use. Hence, mutatis mutandis, the assumption of densely observed data is similar to the one of Kargin and Onatski (2008) and Kokoszka et al. (2015, 2018), among others. Kargin and Onatski (2008) consider a functional autoregressive problem to predict curves, for example the yield curve using the past observed one. They do not constrain their results to a finite number of factors, hence they need to use regularization. Kokoszka et al. (2015) recasts the usual factor model into a functional factor model to estimate the constant factor beta using multiple intraday frequencies and an intercept that is allowed to vary with the time frequency. Kokoszka et al. (2018) considers the sorted cross-section of stock returns on a monthly basis and recast such returns into a functional variable, where the index is the fraction of sorted assets. The estimation in the latter two references does not require any regularization due to the problem structure. More complex estimators exist for sparsely and irregular observed data, but we do not need the extra complication here (e.g. Sancetta, 2015, and references therein).

Within the functional data framework, we derive algorithms for the estimation of functional loadings, risk prices based on linear functional discount factors, and functional risk premia. The algorithms are based on the extension of asset pricing to functional factors. Our econometric results ensure consistency in this setup as well asymptotic normality of the estimators. In particular, the estimators for the functional factors and factor scores are asymptotically Gaussian. However, due to the initial estimation of factor scores, we induce error in variables in the subsequent estimation steps. The covariance of the estimators depend on the covariance of the estimated eigenfunctions. These have a rather complex structure (Bosq, 2000, Ch.4). Hence, we make no attempt to provide the functional form of the covariance of our estimators in the statement of the results. Instead, having shown asymptotic normality, we can rely on the bootstrap to carry out inference.

There is a rich literature on estimation functional regression with scalar response (Horváth and Kokoszka, 2012, Ch.8.4 for references). Our approach corresponds to the standard one where only a few factors scores are assumed to be loaded. This is usually referred to functional principal component regression (Horváth and Kokoszka, 2012, eq. 8.19). More general approaches do not make such assumption and essentially belong to the general family of linear inverse problems (Carrasco et al., 2007 for a treatment of such problems in econometrics, Horváth and Kokoszka, 2012, for the applications to functional regression problems). Once extended to the multivariate case ($L > 1$), these results would be of interest if an infinite number of factor scores were priced. However, for many pricing applications, the number I of priced factor scores would be small. Hence, we focus on the validity of procedures that are more commonly employed by practitioners but justified using functional data analysis.

2.2 Notation

We shall define pricing factors and a linear stochastic discount factor in a functional data context. Given that we need to use functional factors, we provide additional notation to guide the reader through some of the more theoretical results of the following sections.

The reader can skim through the following and use it as reference when reading the statement of the results.

We let $|\cdot|_F$ be the Frobenius norm: $|x|_F^2 = \text{Trace}\,(x'x)$ for any matrix $x \in \mathbb{R}^{u\times v}$. If $v = 1$, this is the Euclidean norm that we denote by $|\cdot|_2$. Let $\mathcal{H}^u = \mathcal{H}^u([0,1])$ be the separable Hilbert space of $u \times 1$ dimensional vector valued functions on $[0,1]$ with inner product $\langle x, y\rangle_{\mathcal{H}^u} = \int_0^1 x(s)' y(s)\, ds$, $x, y \in \mathcal{H}^u$. The norm $|\cdot|_{\mathcal{H}^u}$ is the one induced by the inner product. If $u = 1$, we just write $\mathcal{H} = \mathcal{H}^1$. Let $F_t^{(l)} \in \mathcal{H}$ denotes the l^{th} entry in the $L \times 1$ functional factor $F_t \in \mathcal{H}^L$, and $C_{FF}^{(l)} : [0,1] \times [0,1] \to \mathbb{R}$ be the covariance function of $F^{(l)}$. We denote the k^{th} element in the vector of returns R_t by $R_t^{(k)}$. Throughout, 1_K and 0_K will denote the K-dimensional column vectors of ones and zeros, respectively. While standard, for the sake of clarity, we recall that for any two column random vectors X and Y, we have that $Cov\,(X, Y) = \mathbb{E}\,(X - \mathbb{E}X)\,(Y - \mathbb{E}Y)'$ and $Var\,(X) = Cov\,(X, X)$.

To help the reader's intuition, we consider the following examples.

Example 1 *(CAPM) The CAPM is a special case of our framework. In this case,* $F_t(s) = R_{t,m}^e$ *where* $R_{t,m}^e$ *is the market excess return. Hence, the market excess return is a functional factor that is constant for* $s \in [0,1]$. *Then,* $C_{FF}(s, v) = Cov\,(F_t(s), F_t(v)) = Var\left(R_{t,m}^e\right)$.

Example 2 *(One Functional Factor) Consider a one dimensional functional factor. This corresponds to the case* $F_t \in \mathcal{H}^L$ *with* $L = 1$. *Then,* F_t *is a square integrable function on* $[0,1]$ *and its covariance function* C_{FF} *is real valued. For example, this could be the yield curve with tenors normalised to be in* $[0,1]$.

Example 3 *(Functional and Non-Functional Factors) Our framework enables us to combine scalar and functional factors for inference. This example is relevant to our application in Section 3. Suppose that* $F_t(s) = \left[R_{t,m}^e, \Sigma_t(s)\right]'$, *where* $R_{t,m}^e$ *is the market excess return and* Σ_t *is a functional factor, such as the unpredictable part of the variance smile. This corresponds to the case* $F_t \in \mathcal{H}^L$ *with* $L = 2$. *Then, for* $s \in [0,1]$, *the matrix valued covariance function of the functional factor* F_t *is*

$$C_{FF}(s, s) = \begin{bmatrix} Var\left(R_{t,m}^e\right) & Cov\left(R_{t,m}^e, \Sigma_t(s)\right) \\ Cov\left(R_{t,m}^e, \Sigma_t(s)\right) & Var\left(\Sigma_t(s)\right) \end{bmatrix},$$

and the covariance with the risky assets returns is

$$C_{RF}(s,s) = \left[\begin{array}{cc} Cov\left(R_t, R_{t,m}^e\right) & Cov\left(R_t, \Sigma_t(s)\right)\end{array}\right].$$

Moreover, we define $C_{FF}^{(1)}(s,s) = Var\left(R_{t,m}^e\right)$, $C_{FF}^{(2)}(s,s) = Var\left(\Sigma_t(s)\right)$. *Note that* $C_{FF}^{(1)}(s,s)$ *does not change with* $s \in [0,1]$ *and similarly the first column in* C_{RF}.

2.3 Pricing Factors

The variable F_t is supposed to have a finite second moment and represents L functional factors.[1] Next, we define when F_t is a pricing factor.

Definition 1 *Let* R_t *be the* K*-dimensional column vector of returns with finite second moment at time* t. *Then,* $F_t \in \mathcal{H}^L$ *is a pricing factor for* R_t *if there is a scalar* α *and a* $\lambda \in \mathcal{H}^L$ *such that (1) holds where for any* $\gamma \in \mathcal{H}^L$, *the transpose of the* k^{th} *row of* β, *is* $\beta^{(k)} \in \mathcal{H}^L$ *satisfying*

$$Cov\left(R_t^{(k)} - \int_0^1 \beta^{(k)}(s)' F_t(s)\, ds, \int_0^1 \gamma(s)' F_t(s)\, ds\right) = 0 \tag{2}$$

for $k = 1, 2, ..., K$.

Definition 1 allows us to use functional pricing factors. When $\mathcal{H}^L$ is $\mathbb{R}^L$ and $Var(F_t)$ is invertible, (2) implies that $\beta = Cov(R_t, F_t)\, Var(F_t)^{-1}$, as usual.

The fact that F_t is a functional factor in $\mathcal{H}^L$ has nontrivial implications for $Cov(R_t, F_t)$ as the following result shows.

Lemma 1 *Let the coefficients* $(\theta_i)_{i\geq 1}$ *and the functions* $\left\{\Psi_i \in \mathcal{H}^L : i \geq 1\right\}$ *be the eigenvalues and related eigenfunctions of the* $L \times L$ *matrix valued covariance function* $C_{FF} := \{Cov(F_t(v), F_t(s)) : v, s \in [0,1]\}$, *where* $\theta_i \geq \theta_{i+1}$, $i \geq 1$ *and* $\theta_i \to 0$. *Then, in Definition 1,* $\beta^{(k)} \in \mathcal{H}^L$ *if and only if*

$$\sum_{i=1}^{\infty} \frac{Cov\left(R_t^{(k)}, \langle F_t, \Psi_i\rangle_{\mathcal{H}^L}\right)^2}{\theta_i^2} < \infty \tag{3}$$

uniformly in $k = 1, 2, ..., K$.

[1] With no further mention, all vectors are column vectors.

Lemma 1 says that the correlations of the factor scores $\{\langle F_t, \Psi_i \rangle_{\mathcal{H}^L} : i \geq 1\}$ with the asset returns needs to go to zero fast enough for F_t to be a pricing factor. From a practical point of view, we could think that only the first few factor scores are priced, while the rest are eventually redundant.

2.4 Stochastic Discount Factor

For the absence of arbitrage, we require a stochastic discount factor M_t such that $\mathbb{E}(1_K + R_t) M_t = 1_K$. Let $R_t^{(0)}$ be the return on the zero beta portfolio. Its expectation is denoted by $\alpha := \mathbb{E}R_t^{(0)}$. The return $R_t^{(0)}$ satisfies $Cov\left(R_t^{(0)}, M_t\right) = 0$ by definition. By no arbitrage, $\mathbb{E}\left(1 + R_t^{(0)}\right) M_t = 1$, which can be rewritten as

$$\mathbb{E}\left(1 + R_t^{(0)}\right) M_t = Cov\left(R_t^{(0)}, M_t\right) + \mathbb{E}\left(1 + R_t^{(0)}\right) \mathbb{E}M_t.$$

This implies that $\mathbb{E}M_t = 1/(1+\alpha)$. Under the regularity conditions presented in Lemma 2 below, the linear relation in (1) is equivalent to the existence of a linear stochastic discount factor

$$M_t = \frac{1 - \int_0^1 [(1 - \mathbb{E}) F_t(s)]' b(s)\, ds}{(1+\alpha)}, \tag{4}$$

where $\{b(s) : s \in [0,1]\}$ is an element in $\mathcal{H}^L$. For this to hold, we are assuming that F_t is a pricing factor as in Definition 1.

The above displays together with the previous discussion imply that

$$\mathbb{E}R_t = \alpha 1_K + \int_0^1 Cov(R_t, F_t(s))\, b(s)\, ds. \tag{5}$$

Hence, b is the price of factor covariance risk. We shall refer to it simply as risk price, throughout. We can now extend the standard relation between the risk price b and the risk premium to the functional framework. This is equivalent to showing that (5) is equal to the beta pricing model in (1). Under fast decay in correlation between the returns and the factor scores as in (3), we can show that this is the case if and only if the risk premia satisfy a certain decay condition as stated in the following.

Lemma 2 *Suppose that (3) holds. Then,* $\int_0^1 Cov(R_t, F_t(s))\, b(s)\, ds = \int_0^1 \beta(s) \lambda(s)\, ds$ *for* β *as in Definition 1 and* $b \in \mathcal{H}^L$ *if and only if* $\sum_{i=1}^{\infty} \langle \lambda, \Psi_i \rangle_{\mathcal{H}^L}^2 / \theta_i^2 < \infty$.

In the finite dimensional case, i.e. $F_t(s)$ constant for each $s \in [0,1]$ as in the CAPM Example 1, the lemma is trivial, as we have that $\beta = Cov(R_t, F_t) Var(F_t)^{-1}$, $\lambda = Var(F_t)\, b$. When (3) holds, we still have that the risk price $b(s)$ can be mapped into the risk premium via a linear transformation. However, the transformation has non-trivial implications. From a practical point of view, it states in a different way the conclusion of Lemma 1: the risk premium of factor scores must admit a series expansion $\lambda = \sum_{i=1}^{\infty} \langle \lambda, \Psi_i \rangle_{\mathcal{H}^L} \Psi_i$ where the coefficients $\langle \lambda, \Psi_i \rangle_{\mathcal{H}^L}$ decay fast.[2]

Next we discuss series expansions for the quantities of interest. We can then define estimation procedures for functional factors using these expansions.

2.5 Series Representation for Functional Data Factors and Risk Premia

We need to establish series representations that will allow us to derive approximations and expressions for the estimators of the functional betas, risk prices, and the risk premia. We shall use the following condition.

Condition 1 *The factor $F_t^{(l)}$ is in $\mathcal{H}$, $\mathbb{E}\left|F_t^{(l)}\right|_{\mathcal{H}}^2 < \infty$, and its covariance function is continuous, $l = 1, 2, ..., L$.*

The linear stochastic discount factor (4) together with Condition 1 allows us to derive the series expansions for the quantities that we have introduced. These are the basis for the construction of our estimators. They will also be used in our empirical analysis. Under Condition 1, the series expansion for the factors holds under the uniform norm. Moreover, we focus on expansions for each individual factor separately. Examples will follow at the end of this section.

Theorem 1 *Let R_t be the K-dimensional column vector of returns with finite second moment at time t. Suppose the stochastic discount factor (4) and that Condition 1 holds. We have the following.*

1. *There are L orthonormal bases of $\mathcal{H}$, $\left\{\left\{\Phi_i^{(l)} : i = 1, 2, ...\right\} : l = 1, 2, ..., L\right\}$, such that,*

$$Cov\left(F^{(l)}(s), F^{(l)}(v)\right) = \sum_{i=1}^{\infty} \rho_i^{(l)} \Phi_i^{(l)}(s)\, \Phi_i^{(l)}(v)' \tag{6}$$

[2]The condition on λ means that it must be an element in the reproducing kernel Hilbert space of the kernel κ such that $\kappa(v,s) = \int_0^1 C_{FF}(v,r)\, C_{FF}(r,s)\, dr$. However, this fact is not further exploited in the rest of the paper.

for absolutely summable scalar coefficients $\rho_i^{(l)}$ *satisfying* $\rho_i^{(l)} \geq \rho_{i+1}^{(l)} \geq 0$ *and the series on the right hand side converges uniformly in* $s, v \in [0,1]$.

2. *The factors admit the representation*

$$F_t^{(l)}(s) = \mathbb{E}F_t^{(l)}(s) + \sum_{i=1}^{\infty} \sqrt{\rho_i^{(l)}} \xi_{t,i}^{(l)} \Phi_i^{(l)}(s), \; l = 1, 2, ..., L, \tag{7}$$

where $\xi_t^{(l)} := \left\{ \xi_{t,i}^{(l)} : i = 1, 2, ... \right\}$ *is a sequence of mean zero, variance one uncorrelated random variables, and the second moment of the right hand side converges uniformly in* $s \in [0,1]$.

3. *If (3) holds, the pricing equation (1) with* β *as in (2) holds with*

$$\int_0^1 \beta^{(k)}(s) \lambda(s) \, ds = \sum_{l=1}^{L} \sum_{i=1}^{\infty} \beta_i^{(k,l)} \lambda_i^{(l)}, \; k = 1, 2, ..., K, \tag{8}$$

for some scalar coefficients $\beta_i^{(k,l)}$ *and* $\lambda_i^{(l)}$ *such that* $\sum_{i=1}^{\infty} \left(\left| \beta_i^{(k,l)} \right|^2 + \left| \lambda_i^{(l)} \right|^2 \right) < \infty$, $k = 1, 2, ..., K$, $l = 1, 2, \ldots, L$. *In particular, denoting the* l^{th} *entry in* $\beta^{(k)}, \lambda \in \mathcal{H}^L$ *by* $\beta^{(k,l)}$ *and* $\lambda^{(l)}$, *we have that*

$$\beta^{(k,l)}(s) = \sum_{i=1}^{\infty} \beta_i^{(k,l)} \Phi_i^{(l)}(s), \; \lambda^{(l)}(s) = \sum_{i=1}^{\infty} \lambda_i^{(l)} \Phi_i^{(l)}(s),$$

where equality is under the norm $|\cdot|_{\mathcal{H}}$, $l = 1, 2, ..., L$.

4. *Then, we have that*

$$\int_0^1 Cov(R_t, F_t(s)) \, b(s) \, ds = \sum_{l=1}^{L} \sum_{i=1}^{\infty} \sqrt{\rho_i^{(l)}} \left(\mathbb{E} R_t \xi_{t,i}^{(l)} \right) b_i^{(l)}$$

for scalar coefficients satisfying $\sum_{i=1}^{\infty} \left| b_i^{(l)} \right|^2 < \infty$. *In particular, The* l^{th} *entry in* $b(s) \in \mathcal{H}^L$ *can be written as*

$$b^{(l)}(s) = \sum_{i=1}^{\infty} b_i^{(l)} \Phi_i^{(l)}(s), \tag{9}$$

and the equality in (9) is under the norm $|\cdot|_{\mathcal{H}}$, $l = 1, 2, ..., L$. *Furthermore, the* *pricing equation (5) can be represented as*

$$\mathbb{E}R_t = \alpha 1_K + \sum_{l=1}^{L}\sum_{i=1}^{\infty}\sqrt{\rho_i^{(l)}}\left(\mathbb{E}R_t\xi_{t,i}^{(l)}\right)b_i^{(l)}. \tag{10}$$

To avoid any ambiguity due to the use of the same symbol for elements in $\mathcal{H}$ and the scalars in the series representation, we shall always use a subscript i when referring to the coefficients in the series representation of $\beta^{(k,l)}$, $\lambda^{(l)}$ and $b^{(l)}$.

Theorem 1 is the main representation that allows us to carry out estimation of the functional risk prices and premia. The decompositions in Theorem 1 are stated elementwise for each $l = 1, 2, ..., L$ as we do want to retain the characteristics of each factor by computing factor scores for each $F_t^{(l)}$ separately. Finally, the variables $\xi_t^{(l)}$ are possibly correlated across l and t. Throughout, we shall write $S_{t,i}^{(l)} := \sqrt{\rho_i^{(l)}}\xi_{t,i}^{(l)}$ so that $S_{t,i}^{(l)}$ is a factor score.

For practical application of Theorem 1, we assume that only the first I factor scores are priced for each factor $l = 1, 2, ..., L$. By this, we mean that the risk price (9) admits the representation $b^{(l)}(s) = \sum_{i=1}^{I} b_i^{(l)}\Phi_i^{(l)}(s)$, $l = 1, 2, ..., L$.

The next result establishes the linear beta decomposition of excess returns in terms of risk exposure to factors scores plus an orthogonal component which is not priced. This is the extension of a standard result that holds in the less familiar functional factors framework.

Lemma 3 *Suppose that Condition 1 holds and that for each functional factor only the first I factor scores are priced.*

1. *Then, (5) reduces to $\mathbb{E}R_t^e = Cov(R_t, S_t)\, b_0$ where $R_t^e = R_t - 1_K R_t^{(0)}$, the return in excess of the zero beta portfolio, and S_t and b_0 are $LI \times 1$ vectors with $i + (l-1) I$ entry equal to $S_{t,i}^{(l)}$ and $b_i^{(l)}$, respectively.*

2. *If the factor scores have full rank covariance matrix, this also implies that $R_t^e = a + BS_t + \varepsilon_t$, where $a = \mathbb{E}R_t^e - B\mathbb{E}S_t$ is $K \times 1$, $B = Cov(R_t, S_t)\, Var(S_t)^{-1}$ is $K \times LI$ and ε_t is $K \times 1$ vector uncorrelated with S_t. If the factor scores S_t are also tradable, $a = 0_K$.*

3. *Finally, we also have that $\mathbb{E}R_t^e = B\Lambda$ where Λ is an $LI \times 1$ vector with $i + I(l-1)$ entry equal to $\lambda_i^{(l)}$ as in (8).*

The indexing in Lemma 3 is based on the following convention. The factor score S_t is obtained stacking the vectors $\left(S_{t,1}^{(l)}, ..., S_{t,I}^{(l)}\right)'$ one below the other, $l = 1, 2, ..., L$. A similar comment applies to the other quantities. This ordering convention is followed throughout.

We shall build on Examples 1, 2 and 3 to clarify the current framework.

Example 4 *(CAPM, Example 1 Cont'd) Recall that $F_t(s) = R_{t,m}^e$ is the market excess return. Trivially, the CAPM implies that we only have one eigenvalue $\rho_1 = Var\left(R_{t,m}^e\right)$ with eigenfunction $\Phi_1(s) = 1$ while $\rho_i = 0$ for $i \geq 2$. Then, the first factor score is $S_{t,1} = R_{t,m}^e - \mathbb{E}R_{t,m}^e$, while $S_{t,i} = 0$ for $i \geq 2$. Using the notation in Theorem 1, $S_{t,1} = \sqrt{\rho_1}\xi_{t,1}$. Moreover, $Cov(R_t, F_t(s)) = \mathbb{E}\left[R_t\left(R_{t,m}^e - \mathbb{E}R_{t,m}^e\right)'\right]$. Hence, we recover the usual results for the CAPM: $\mathbb{E}R_t^e = \beta_m\lambda_m$, where β_m is the beta on the market risk premium λ_m.*

Example 5 *(One Functional Factor Only, Example 2 Cont'd) Without change in notation, this is the setup of Theorem 1 with $L = 1$. Then, $C_{FF}(s, v) = Cov(F_t(s), F_t(v)) = \sum_{i=1}^{\infty} \rho_i \Phi_i(s) \Phi_i(v)$. Suppose that there K non-redundant assets priced in the economy. From Theorem 1, we have that*

$$\mathbb{E}R_t = \alpha 1_K + \sum_{i=1}^{\infty} \sqrt{\rho_i}\left(\mathbb{E}R_t\xi_{t,i}\right) b_i.$$

Suppose that only the first I price of risk coefficients b_i are non-zero (we are writing $b_i = b_i^{(1)}$ because $L = 1$). Using Lemma 3, the above display can be written as $\mathbb{E}R_t^e = B\Lambda$. Note that $\mathbb{E}R_t^e$ is in the span of K linearly independent vectors in $\mathbb{R}^K$. Hence, we can identify the risk premium on at most K factor scores. Moreover, if the functional factor is tradable (i.e. factor scores are tradable), we must have $I < K$. If $I = K - 1$, then all K risky assets are redundant, i.e. they are a linear combination of the $K - 1$ tradable factor scores and the zero beta portfolio.

Example 6 *(Functional and Non-Functional Factors, Example 3 Cont'd) The eigenfunctions $\left\{\Phi_i^{(l)} : i = 1, 2, ..., I\right\}$ are derived from $\left\{C_{FF}^{(l)}(s, s) : s \in [0, 1]\right\}$ $l = 1, 2$. For $l = 1$ these are just one for $i = 1$ and zero otherwise because $C_{FF}^{(1)}(s, s)$ does not depend on s, as in Example 4. Under the assumption that only the first I factor scores are priced, we can rewrite $\mathbb{E}R_t = \alpha + \beta_m\lambda_m + \int_0^1 \beta_\Sigma(s)\lambda_\Sigma(s)\,ds$ as $\mathbb{E}R_t = B\Lambda$ where $B = [1_K, \beta_m, \beta_{\Sigma,1}, ..., \beta_{\Sigma,I}]$ and $\Lambda = [\alpha, \lambda_m, \lambda_{\Sigma,1}, ..., \lambda_{\Sigma,I}]'$. Here, the vectors β_m and*

$\beta_{\Sigma,i}$ are $K \times 1$. They are the loadings on the market and on the i^{th} factor scores of Σ. On the other hand, λ_m and $\lambda_{\Sigma,i}$ are the associated scalar risk premia. Making α a parameter to estimate allows us to test whether the pricing restriction holds. When we use the risk free interest rate for the zero beta portfolio, α is interpreted as the risk free interest rate. Recall that $R_t^e = R_t - R_t^{(0)} 1_K$, then in $\mathbb{E} R_t^e = B\Lambda$, the first entry in Λ is the pricing error and this should be zero. Recall that $R_t^{(0)}$ is the return on the zero beta portfolio.

We now focus on estimators for the above quantities as outlined in the following section.

2.6 Estimation Algorithms

We consider a two-step estimation of loadings and risk premia and the generalised method of moments estimator of the risk prices b. In what follows we use $\hat{\mathbb{E}}$ to denote empirical expectation over the time index t, i.e. the sample average, based on a sample of n observations.

The asymptotic properties of the estimators will be studied in Section 2.7. Finite sample properties are also studied via simulations in Section A.4 of the Supplementary Material.

2.6.1 Algorithm for Functional Two-Step Procedure

The classical two-step procedure consists of a time series regression to estimate the factor loadings and a cross-sectional regression of the time averaged returns on the estimated loading to find the risk premia. In the context of functional factors the procedure is similar and is shown in Algorithm 1. The loadings are directly estimated relying on the restrictions imposed by Definition 1.

Algorithm 1 Two-Step Estimation.

1. Estimate $\hat{C}_{FF}^{(l)}$ as sample counterpart of $C_{FF}^{(l)}$ and find its first I empirical eigenfunctions $\left\{\hat{\Phi}_i^{(l)} : i = 1, 2..., I\right\}$, $l = 1, 2, .., L$.
2. Compute the sample factor scores $\hat{S}_{t,i}^{(l)} = \left\langle F_t^{(l)}, \hat{\Phi}_i^{(l)} \right\rangle_{\mathcal{H}}$, $\forall t, i, l$ and define $\hat{S}$ to be the $n \times LI$ matrix with $(t, i + (l-1) I)$ entry equal to $\hat{S}_{t,i}^{(l)}$ and $\hat{S}_t$ the transpose of the t row of $\hat{S}$.
3. Compute the $LI \times LI$ matrix $\hat{C}_{\hat{S}\hat{S}} = \hat{\mathbb{E}}\left(\hat{S}_t - \hat{\mathbb{E}}\hat{S}_t\right)\left(\hat{S}_t - \hat{\mathbb{E}}\hat{S}_t\right)'$ and the $K \times LI$ matrix $\hat{C}_{R\hat{S}} = \hat{\mathbb{E}} R_t \left(\hat{S}_t - \hat{\mathbb{E}}\hat{S}_t\right)'$.
4. Compute the $K \times LI$ matrix $\hat{B} = \hat{C}_{R\hat{S}} \hat{C}_{\hat{S}\hat{S}}^{-1}$.
5. Estimate the $LI \times 1$ vector $\hat{\Lambda} = \left(\hat{B}'\hat{B}\right)^{-1} \hat{B}'\hat{\mathbb{E}} R_t$ where the $i + (l-1) I$ entry is an estimator for $\lambda_i^{(l)}$ in Theorem 1.

In Point 2 of Algorithm 1, the matrix $\hat{S}$ is constructed appending the I factor scores for each factor l one on the right of each other. This is the same ordering convention discussed right after Lemma 3.

While we write the algorithm assuming continuous argument $s \in [0, 1]$ in the functional factors and covariances, in practice we only sample the factors at a discrete set of points $\mathcal{S}_N := \{s_1, s_2, .., s_N\} \subset [0, 1]$. This means, that integrals are replaced by averages over N terms. In this case, the eigenfunctions are approximated by $\sqrt{N}$ the eigenvectors of the covariance matrix $\left\{C_{FF}^{(l)}(s, t) : s, t \in \mathcal{S}_N\right\}$ and the eigenvalues by N^{-1} times the corresponding matrix eigenvalues. Note that these adjustments ensures that the resulting quantities converge to the true values and the scaling does not change with N (e.g. Rasmussen and Williams, 2006, p.99). This is because the matrix eigenvalues grow linearly with N and the matrix eigenvectors decrease as $N^{-1/2}$. In practice, given that N is fixed, we may just use the eigenvectors and related eigenvalues if consistent scaling is not required.

2.6.2 Algorithm for Functional Discount Factor Estimation

Let $R_t^{(k)}$ be the k^{th} entry in R_t . We consider the GMM estimator:

$$\frac{1}{K} \sum_{k,l=1}^{K} \hat{W}_{k,l} \left(\frac{1}{n} \sum_{t=1}^{n} M_t(b) R_t^{(k)}\right) \left(\frac{1}{n} \sum_{t=1}^{n} M_t(b) R_t^{(l)}\right), \tag{11}$$

where $M_t(b)$ is the candidate discount factor (as in (4)), which depends on an unknown parameter $b \in \mathcal{H}^L$, and $\hat{W}_{k,l}$ is the k, l entry of a possibly estimated matrix $\hat{W}$. When $\hat{W}$ is the inverse of the sample second moment matrix of the returns, (11) becomes a sample estimator of the Hansen-Jagannathan distance. The estimator $\hat{W}$ can be noisy, hence it might be preferable to restrict it to be diagonal. For this reason, we shall constrain $\hat{W}$ to be diagonal in the empirical application in Section 3. Given that diagonal $\hat{W}$ also simplifies the technical arguments, we shall assume it when proving consistency and normality of the estimator in Section 2.7.2.[3] Following the notation from Theorem 1, we assume that $b^{(l)}(s) = \sum_{i=1}^{I} b_i^{(l)} \Phi_i^{(l)}(s)\ s \in [0,1]$. This means that only the first I factor scores of $F_t^{(l)}$ are priced for each $l = 1, 2..., L$. Then, Algorithm 2 shows how to compute the estimator $\hat{b}_0$ for b_0 where the latter is as in Lemma 3. Recall that b_0 is the vector that collects the coefficients in the series expansion of $b^{(l)}(s)$, $l = 1, 2, ..., L$.

Algorithm 2 Discount Factor Estimation.

1. Estimate $\hat{C}_{R\hat{S}}$ as in Algorithm 1.
2. Estimate the $LI \times 1$ vector $\hat{b}_0 = \left(\hat{C}'_{R\hat{S}} \hat{W} \hat{C}_{R\hat{S}}\right)^{-1} \hat{C}'_{R\hat{S}} \hat{W} \hat{\mathbb{E}} R_t$ with diagonal $\hat{W}$.

2.7 Asymptotic Analysis

This section establishes the asymptotic properties of the estimators obtained from Algorithms 1 and 2. The finite sample properties of the estimators are studied in a set of simulations in Section A.4 of the Supplementary Material. Conclusions from such simulations are in line with the asymptotic theory established in this section.

We start introducing a set of conditions followed by a brief set of remarks. We then show that the estimators in Algorithms 1 and 2 are consistent and asymptotically normal.

2.7.1 Regularity Conditions

The following regularity conditions will be used in all the results in order to show consistency of the estimators in Algorithms 1 and 2, as well as their weak convergence to Gaussian random elements after proper scaling and centering.

[3] At the cost of technical complexity and additional moment conditions, our asymptotic results would hold if $\hat{W}/K$ had square summable entries for $K \to \infty$.

Condition 2 *The factor process* $\{F_t : t = 1, 2, ...\}$ *is a sequence of i.i.d. mean zero random variables satisfying Condition 1, and* $\mathbb{E}\left|F_t\right|^4_{\mathcal{H}^L} < \infty$.

Condition 3 *The returns* $\{R_t : t = 1, 2, ...\}$ *are i.i.d. random variables with values in* $\mathbb{R}^K$ *such that* $\max_{k \leq K} \mathbb{E}\left|R_t^{(k)}\right|^4 < \infty$ *(*$R_t^{(k)}$ *is the* k^{th} *entry in* R_t*), where* K *is a positive integer, possibly growing with the sample size. The same distributional and moment conditions apply to the zero beta portfolio.*

Condition 4 *Let* $C_{FF}^{(l)} := \left\{Cov\left(F_t^{(l)}(r), F_t^{(l)}(s)\right) : r, s \in [0, 1]\right\}$. *The first* $I + 1$ *eigenvalues of* $C_{FF}^{(l)}$ *are distinct,* $l = 1, 2, ..., L$.

Condition 5 *Let* S_t *be the* $LI \times 1$ *vector with* $i + (l-1) I$ *entry equal to the factor score* $S_{t,i}^{(l)}$*. The* $LI \times LI$ *matrix* $C_{SS} = Var(S_t)$ *and the* $K \times LI$ *matrix* $C_{RS}/K^{1/2} = Cov(R_t, S_t)/K^{1/2}$ *have singular values contained in a compact interval inside* $(0, \infty)$*. For* $C_{RS}/K^{1/2}$ *this holds uniformly in* K*.*

Condition 6 *There is a* $K \times K$ *diagonal matrix* W *such that* $\left|\hat{W} - W\right|_F^2 = O_p\left(Kn^{-1}\right)$*. Moreover,* W *has entries in a compact interval inside* $(0, \infty)$*, which is independent of* K*.*

We shall refer to the above as the Regularity Conditions. In what follows we implicitly assume that L and I are finite and fixed integers. On the other hand, we do not restrict the number of risky assets K.

Next we remark on the conditions. When dealing with pricing models, we can consider the unpredictable part of the quantities of interest. Hence there is no loss of generality to assume F_t to have mean zero. The i.i.d. assumption is used to avoid distracting technicalities. It is possible to account for dependence at the cost of additional technicalities (Bosq, 2000, Horváth and Kokoszka, 2012, Ch. 16). Conditions 5 and 6 are only used for analysis of the GMM estimator. Regarding Condition 5, if only the first I factor scores are priced for each factor, Lemma 3 implies that $C_{RS} = BVar(S_t)$. The smallest singular value is the square root of the smallest eigenvalue of $C'_{RS}C_{RS} = Var(S_t) B'BVar(S_t)$. Given that $Var(S_t)$ is full rank by assumption, the singular values of C_{RS} are proportional to the ones of B. Then, Condition 5 means that $B'B$ must have eigenvalues that grow linearly with K. This is reasonable as each entry in $B'B$ is the sum of K elements. Clearly, this is true for the CAPM. To simplify the notation, we have not made I dependent on the factor $F_t^{(l)}$, $l = 1, 2, ..., L$.

Condition 6 says that the estimated scaling matrix with dimension $K \times K$ is consistent in the Frobenius norm divided by $K^{1/2}$. In the empirical study, we shall suppose that $\hat{W}$ is proportional to the inverse of the sample variance of the returns. This choice makes pricing implied by (11) invariant to returns variability keeping the focus on the original portfolios because the scaling matrix is diagonal.

2.7.2 Consistency and Asymptotic Normality of the Estimators

In this section we show consistency and asymptotic normality of the estimators obtained from Algorithms 1 and 2.

In the following results, the Regularity Conditions (Conditions 2, 3, 4) are tacitly supposed to hold. We shall use, without mention, the notation in these regularity conditions and in Theorem 1 and Lemma 3. The central limit theorem (CLT) results are stated without explicitly showing the form of the covariance of the process. This is because the expressions are too complex for practical use, but details are given in the proofs. The CLT results justify the use of the bootstrap, which is the natural route to conduct inference in the present context.

Theorem 2 *Let A be a $K \times p$ matrix with rank p, where p is fixed, and such that $\mathbb{E} \left| A' R_t \right|_2^4 < \infty$. For $B = C_{RS} C_{SS}^{-1}$, using Algorithm 1, $\sqrt{n} A' \left(\hat{B} - B \right) \to G_A$ in distribution, where G_A is a mean zero $p \times LI$ Gaussian random matrix. We also have that $K^{-1/2} \left| \hat{B} - B \right|_F \to 0$. These results hold true even when $K \to \infty$, if $\limsup_K \mathbb{E} \left| A' R_t \right|_2^4 < \infty$.*

The matrix A ensures that we can establish asymptotic normality of quantities whose dimension K goes to infinity. This is a common approach in high dimensional econometrics (e.g. Li et al., 2015, Theorem 4.3). The matrix can be used to pick up a finite number of elements in $\hat{B}$. More generally, it can pick up all the elements in $\hat{B}$ but in a constrained way. Hence, its role is equivalent to testing p restrictions in the true B. One application of Theorem 2 is to test whether an additional factor score is loaded. This can be instrumental in choosing the number of factor scores I. In what follows we shall show convergence to Λ and b_0 as defined in Lemma 3.

Theorem 3 *Suppose that for each factor only the first I factor scores are priced. Using Algorithm 1, $\sqrt{n} \left(\hat{\Lambda} - \Lambda \right) \to G_\Lambda$ in distribution, where G_Λ is an $LI \times 1$ mean zero Gaussian vector. We also have that $\left| \hat{\Lambda} - \Lambda \right|_2 \to 0$ in probability. The result holds even when $K \to \infty$.*

Theorem 4 *Suppose that for each factor only the first I factor scores are priced. Let b_0 be as in Lemma 3. Using Algorithm 2, $\sqrt{n}\left(\hat{b}_0 - b_0\right) \to G_{b_0}$ in distribution, where G_{b_0} is a Gaussian vector with mean zero. The result holds even when $K \to \infty$.*

Despite some of the results being stated in terms of loadings of estimated eigenfunctions, we can recover the functional parameters and show that they converge to a Gaussian process.

Lemma 4 *Suppose that $\hat{a} := \left\{\hat{a}_i^{(l)} : l = 1, 2, ..., L; i = 1, 2, ..., I\right\}$ is an $LI \times 1$ random vector such that, for some $a \in \mathbb{R}^{LI}$ $\sqrt{n}\,(\hat{a} - a)$ converges in distribution to a Gaussian vector with mean zero. Let $G_n \in \mathcal{H}^L$ be such that its l entry is $\sqrt{n}\left(\sum_{i=1}^{I} \hat{a}_i^{(l)} \hat{\Phi}_i^{(l)} - \sum_{i=1}^{I} a_i^{(l)} \Phi_i^{(l)}\right)$. Then, G_n converges weakly to a mean zero Gaussian process in $\mathcal{H}^L$ with continuous sample paths.*

Lemma 4 can be applied to Theorems 2, 3 and 4. For example, we can define $\hat{\lambda}^{(l)} = \sum_{i=1}^{I} \hat{\lambda}_i^{(l)} \hat{\Phi}_i^{(l)}$ and show that $\sqrt{n}\left(\hat{\lambda}^{(l)} - \lambda^{(l)}\right)$ converges to a Gaussian process in $\mathcal{H}$ where $\lambda^{(l)}$ is a in Theorem 1. Here, we are denoting by $\hat{\lambda}_i^{(l)}$ the $i + (l-1)\, I$ entry in $\hat{\Lambda}$. Estimation of the factor scores makes the covariance of the limiting Gaussian processes complex for all of the above results. For this reason, we suggest to use the bootstrap. Validity of the bootstrap follows from the asymptotic normality of the estimators.

3 Application: Momentum and the Variance Smile as its Pricing Factor

To show the scope of our approach empirically, we consider empirical asset pricing of momentum sorted portfolios using the market excess returns and the S&P500 implied volatility curve.

3.1 Motivation

Momentum strategies cannot be reconciled with the CAPM and the Fama and French (1993) three-factor model, and their payoff does not necessarily appear to be linear with the market (Moskowitz et al., 2012, Daniel and Moskowitz, 2016). It was observed that a higher level of uncertainty is associated with higher performance of momentum strategies (Hong et al., 2000, Zhang, 2006). However, the literature also showed that momentum is negatively affected by volatility. Its performance can be considerably increased by reducing risk exposure in periods of high volatility (Barroso and Santa-Clara, 2015, Daniel and Moskowitz, 2016). We use the implied volatility curve as a functional factor to shed further light on this anomaly. Our theory justifies the econometric analysis for this problem.

We apply our methodology using the market and the implied variance smile. Our main finding is that curvature/convexity is priced in momentum sorted portfolios. The curvature often increases when the level decreases as shown in Figure 1, where we observe no average convexity during the financial crisis of 2008 as opposed to bull years such as 2013 and 2017. This means that curvature is associated to good states of the world. Hence, holding curvature pays when we transition to such states. In consequence, convexity will have to earn a positive risk premium. The empirical results show that the loading on convexity of momentum portfolios is positive. Hence, such portfolios earn a positive risk premium. In our sample, find that the up minus down (UMD) portfolio (aka momentum portfolio), which is nearly market neutral, has a positive loading on convexity.

3.2 Implied Variance Innovation Process

Let $\mathrm{BS}\left(P_t, \sigma_t, m, T-t\right)$ be the Black and Scholes (B-S) formula for a call at time t on an asset with price P_t, implied volatility σ_t, moneyness m and time to expiry $T-t$. Here, the moneyness m is defined as strike price divided by P_t. In particular,

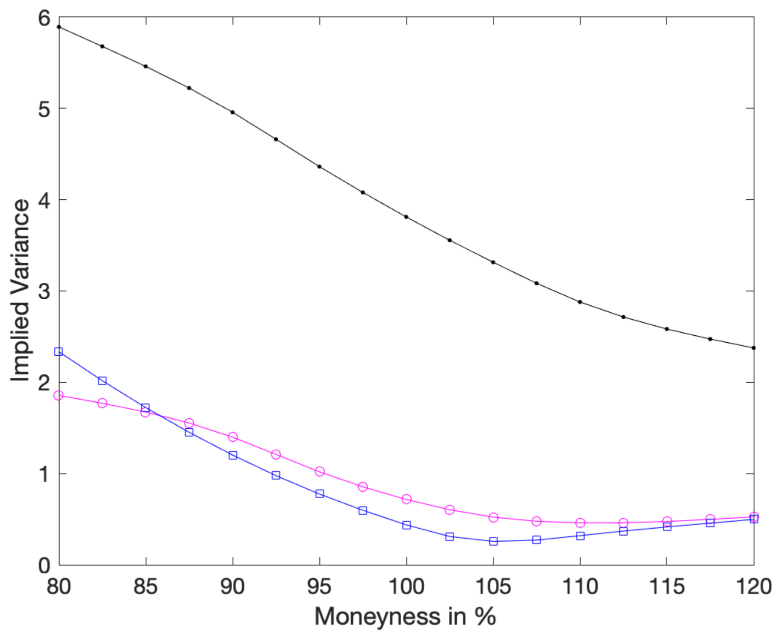
$$\mathrm{BS}\left(P_t, \sigma_t, m, T-t\right) = P_t\left[N\left(d_+\right) - mN\left(d_-\right)\right]$$

where $N(x) = \Pr(Z \leq x)$ and Z is a standard normal random variable and

$$d_\pm = -\frac{\ln(m)}{\sigma_t\sqrt{T-t}} \pm \frac{\sigma_t\sqrt{T-t}}{2}. \tag{12}$$

The implied volatility is the value of volatility that equates the B-S formula to the observed option market price for each given moneyness. We shall work with $\sigma_t^2 = \left\{\sigma_t^2(m) : m \in [\underline{m}, \overline{m}]\right\}$, the variance smile at time t on a 3-month option. In our study $\underline{m} = 0.8$ and $\overline{m} = 1.20$. The smile process is $\left\{\sigma_t^2 : t = 1, 2, 3, ...\right\}$, where time shall be measured at daily frequency. The implied variance process is persistent over time

Figure 1: Implied Variance Curves for SPX. Yearly average of 3-month daily implied variance curves: '.' line is for 2008 (market return: -15.5 basis points (bps) per day), the squares line is for 2013 (market return: +12.2bps per day), the 'o' line is for 2017 (market return: +7.9bps per day). The implied curves have been standardized to daily, dividing them by 252.



t.[4] Let $\mathbb{E}_{t-1}$ be expectation conditioning on information up to time $t-1$. We define the one period innovation to be $\Sigma_t = \sigma_t^2 - \mathbb{E}_{t-1}\sigma_t^2$. For ease of notation, we suppose that the moneyness m has been mapped to $[0,1]$ so that $\{\Sigma_t(s) : s \in [0,1]\}$ where $s = s(m) = (m - \underline{m}) / (\overline{m} - \underline{m})$. Differencing removes most of the time series dependence. This is expected because the implied variance curve should behave locally as a functional random walk plus a drift, which represents a roll yield. In a frictionless market, predictability of the implied variance beyond a risk premium would imply the possibility of arbitrage, as it is a tradable instrument via options.

3.3 Data Description

We use 3-month option implied volatility data on the S&P500 (SPX). Quarterly options are the most traded and this has been traditionally so since the introduction of options

[4]Descriptive statistics in support of this and other related remarks can be found in Section A.3.2 of the Supplementary Material.

on SPX in 1983.[5] These options are listed on the CBOE. The data are obtained from Bloomberg, and their values are expressed as percentages in annualized terms. Each day, the curve is built by Bloomberg calibrating to a lognormal mixture model and backing up the implied vol from the model.[6] This ensures smooth curves and data in the form of functional data as shown in Figure 1. Bloomberg records the volatility smile at moneyness in $\{80, 90, 95, 97.5, 100, 102.5, 105, 110, 120\}$. We interpolate by cubic spline smoothing to ensure equally spaced points with 2.5% moneyness between points. We then map these values into $[0, 1]$ by linear transformation. Hence, our functional observations are actually 17-dimensional vectors. We consider daily frequency for the period 2006/12/27-2019/02/28. We start at the end of 2006, as the dataset from Bloomberg starts in March 2006, but there are missing data. Starting at the end of 2006 ensures that we can construct implied variance differences from January 2007. The sample size is $n = 3062$. For the price data, we use the publicly available data from Kenneth French's data library: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Research`. In particular we use the 10 portfolios sorted by momentum and the up minus down (UMD) portfolio. Both are at daily frequencies.

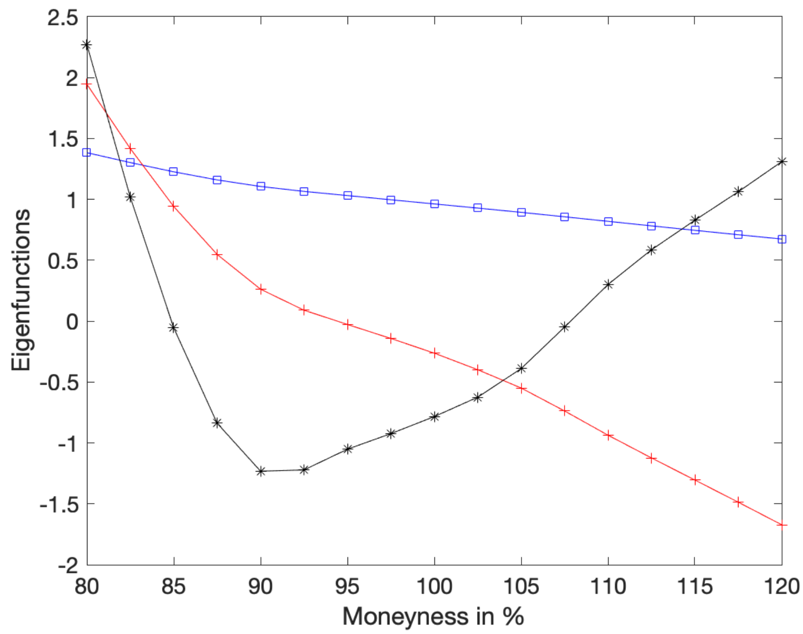
3.4 Empirical Results

The 10 momentum portfolios are sorted from low to high momentum. We also include UMD as 11th portfolio, which is the difference of the highest and lowest momentum portfolio. Although more commonly used as factor, we do include UMD as portfolio to extend available base assets and to understand how such "market neutral" portfolio is priced. We proceed and compute excess returns for all 11 portfolios. The 2008 crisis is characterized by a pronounced increase in the level of variance. After this the level of variance has decreased progressively during the subsequent bull market. An unexpected decrease in the curvature of implied variance appears to be associated with relatively low market performance. We compute the factor scores for the implied variance innovation process Σ. This entails calculation of the empirical eigenvectors of its 17×17 dimensional covariance matrix. We retain the first 3 factor scores which explain about 99% of the total variation. Figure 2 displays the first three eigenvectors. With no loss of generality, these have been signed so that the first value is positive.

[5] http://www.cboe.com/blogs/options-hub/2018/07/02/35-years-of-s-p-500-index-options-trading-at-cboe

[6] See Section A.3.1 in the Supplementary Material for more details.

Figure 2: Eigenfunctions of the smile process. The first three estimated eigenfunctions are plotted. These correspond to level ("squares" line), slope ("+" line) and curvature ("*" line).



From the shape of the eigenvectors, we can interpret the first three factor scores as level slope and curvature. The correlation of these three factor scores with the market is time varying, but strongly negative for the level, mildly negative for the slope, and mildly positive for the curvature.[7]

3.4.1 Preliminary Analysis using Factor Mimicking Portfolios.

Having estimated the factor scores, we compute factor mimicking portfolios for the level, slope and curvature of the implied variance curve.[8] Figure 3 shows the cumulative returns from the factor mimicking portfolios as well as UMD. We note that the level earns negative risk premium as expected. Results for the slope are inconclusive.[9] The curvature earns a positive risk premium. UMD, the winners minus losers portfolio can

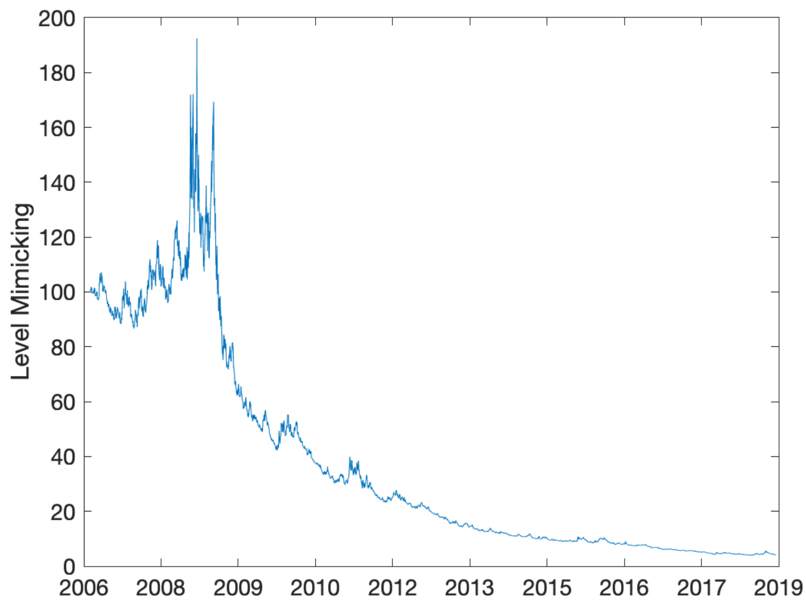
[7]See Section A.3.3 of the Supplementary Material for the details regarding the above remarks.

[8]These factor scores are real valued and the estimation of factor mimicking portfolios is standard. Details and empirical results can be found in Section A.3.6 of the Supplementary Material.

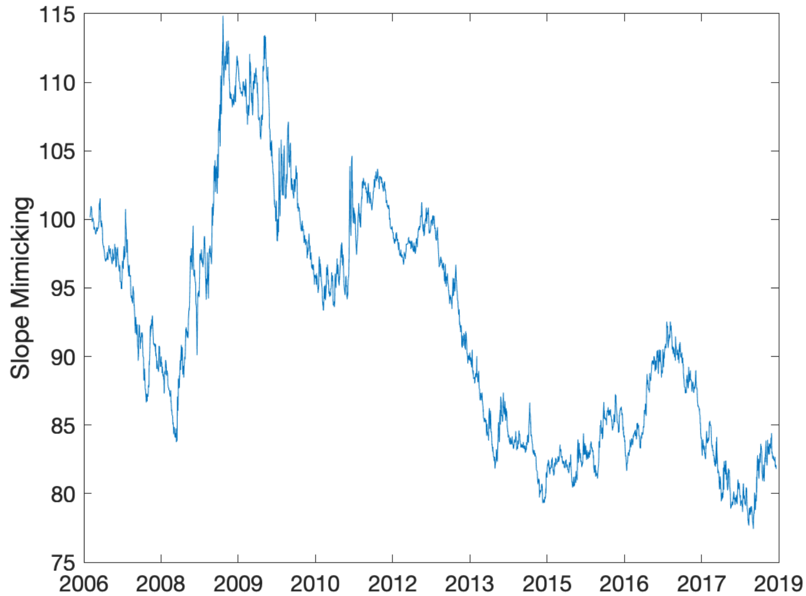
[9]The slope factor score derived here is orthogonal to the level. Hence our result is consistent with the results in Kozhan et al. (2013) who find an insignificant risk premium once a skew swap is hedged by a variance swap.

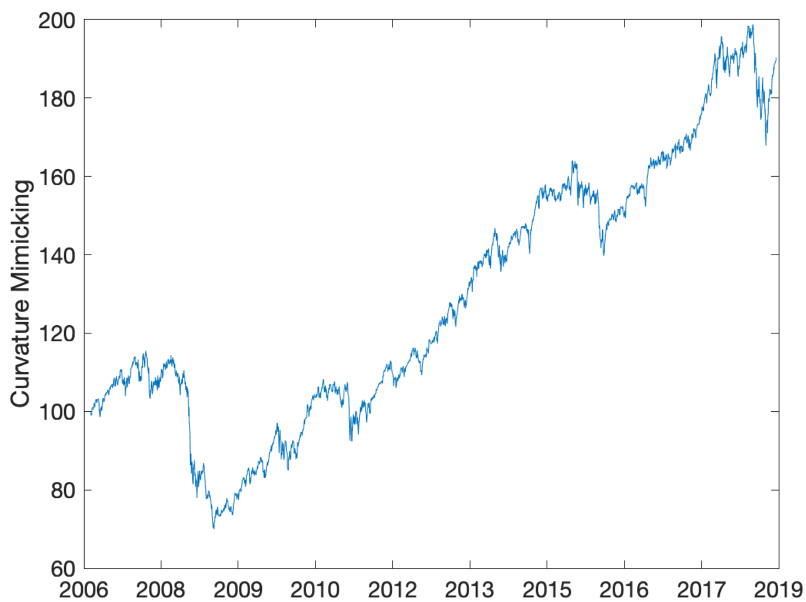
earn a positive risk premium, but is highly exposed to crashes (Barroso and Santa-Clara, 2015, Daniel and Moskowitz, 2016). These results motivate our investigation using an empirical pricing model in the next section. As robustness check we also used monthly data and found qualitatively similar results.[10]

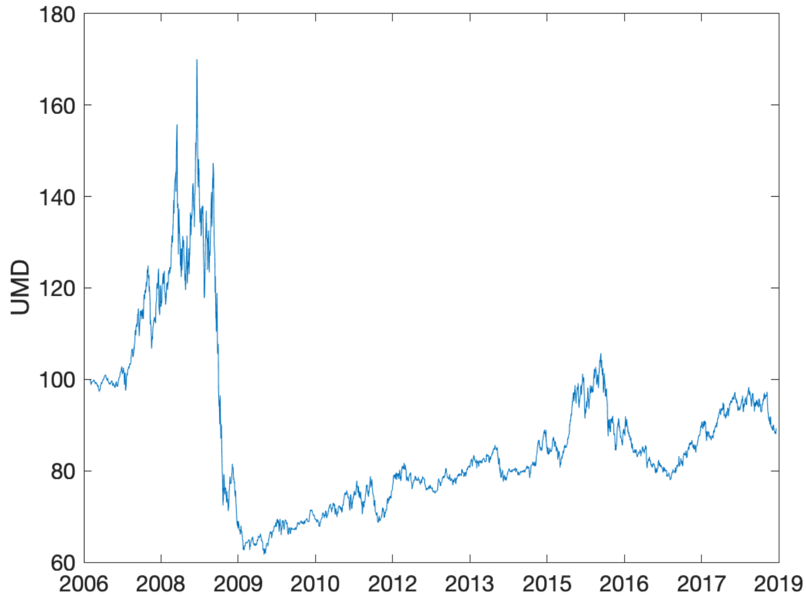
Figure 3: Compounded Returns of Scaled Mimicking Portfolios for Level, Slope, Curvature, and the UMD Factor. The scaling is performed so that the Euclidean norm of the loadings is one.



[10]See Section A.3.6 of the Supplementary Material for a discussion.


Slope Mimicking
115
110
105
100
95
90
85
80
75
2006
2008
2009
2010
2012
2013
2015
2016
2017
2019


Curvature Mimicking
200
180
160
140
120
100
80
60
2006
2008
2009
2010
2012
2013
2015
2016
2017
2019



3.4.2 Estimation Results using Functional Empirical Pricing

We use Algorithms 1 and 2 to carry out our estimations.

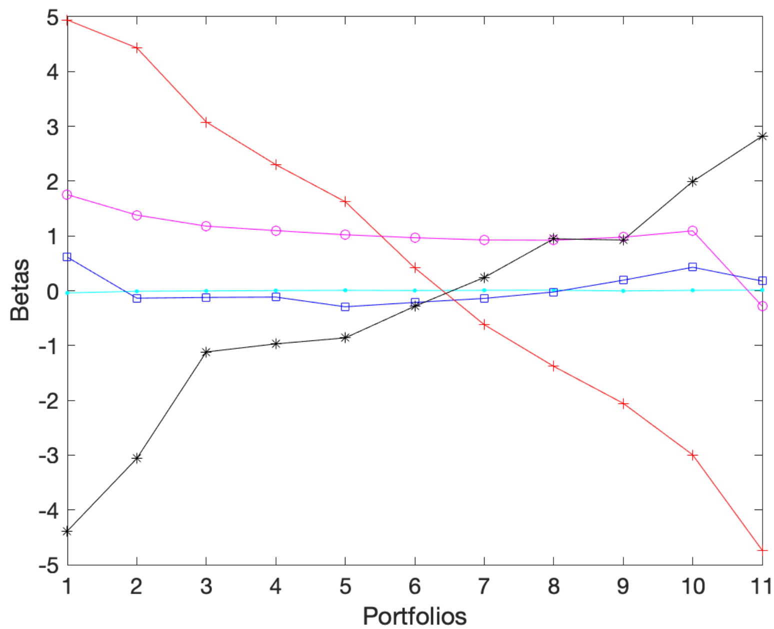
Two-Step Regression Results. Figure 4 plots the estimated loadings $\hat{B}$ from the time series regression and shows a remarkable regularity. Except for the alpha (the intercept) and the level, these estimates of the loadings are significantly different from zero at any conventional level when using bootstrap standard errors.[11] Interestingly, we note that the slope has a negative loading and curvature has a positive loading for high deciles momentum sorted portfolios and for the momentum factor UMD (portfolio 11 in Figure 4).

The cross-section regression of time averaged returns on these loadings produce the estimated risk premia. The results are in Table 1, where we report the non-parametric 95% bootstrap confidence interval.

As often the case, due to the high level of noise in financial data, the confidence interval does contain the origin. However, they are evidently positively shifted in the case of the market and curvature, and negatively shifted in the case of the level. The results are in agreement with the discussion regarding the factor mimicking portfolios. Again, we note that the risk premium on curvature tend to be positive. This result is

[11]Details can be found in Table A.4, Section A.3.4 of the Supplementary Material.

Figure 4: Factor Betas. The alpha ("." line), the market ("o" line), the level ("squares" line), the slope ("+" line) and the curvature ("*" line) are plotted against the portfolios.



new and in contrast to level. Level and curvature are orthogonal components in sample. An unexpected increase in level is associated with transitions into negative states of the world where marginal utility is high. In consequence, under risk aversion, we are willing to pay for insurance in these cases. On the other hand, an unexpected increase in curvature appears to be related to transitions to good states of the world.

Functional Loadings and Premia. Table 2 reports the reconstructed functional loadings and the functional risk premium using the level, slope and curvature factor scores and the estimated eigenfunctions. These are obtained from the estimated eigenfunctions and the factor loadings and risk premia estimates (mutatis mutandis as in Lemma 4). The functional risk premium is negative as expected, but convex in moneyness. For portfolios with highest momentum (i.e. Portfolio 10) as well as for UMD (Portfolio 11), the loadings for high moneyness are larger in absolute value than the loadings for low moneyness. In consequence, a simultaneous and equal increase in high and low moneyness variance (i.e. an increase in curvature) would benefit high momentum portfolios (consistently with the results in Table 1).

Table 1: Cross-Section Regression. The average return of the 11 portfolios are regressed on a constant pricing error and the estimated betas from the time series regression. The bootstrap 2.5% (Q2.5%) and 97.5% (Q97.5%) quantiles are also reported. These represent a 95% bootstrap confidence interval. Numbers are multiplied by 100 and represent basis points per day for the risk premia.

	Pricing Error	Market	Level	Slope	Curvature
Risk Premia	0.0	3.9	-2.6	0.1	0.8
Q2.5%	-0.9	-0.6	-7.1	-1.0	-0.5
Q97.5%	0.8	8.5	0.6	0.9	1.7
R^2	94.7				

Stochastic Discount Factor Estimation. Using Algorithm 2 we estimate the discount factor and report the results in Table 3. These are consistent with the estimates in Table 2 for the two-step regression procedure. The R^2 from these estimations is in excess of 95%.

3.5 Discussion

The results suggest that curvature is a possible driver in the performance of momentum portfolios. To put our results into context, in the supplement to this paper, we compare our results to Carhart four-factor model. Our model fares as well as the later for the sample period. Given the both models use four factors and Carhart model is specifically designed to account for momentum, the results are encouraging.

We did carry out some robustness checks. For example, a large shift in curvature is observable from 2008 to later periods (see Figure 1). Hence, we checked whether we can obtain the same results excluding 2008. By doing so, we found that the loadings of curvature across portfolios became flatter relative to what we observe in Figure 4. Nevertheless, high momentum portfolios still loaded positively the curvature factor score. The estimated risk premium of the curvature factor score was close to zero relative to the results obtained using the full sample. On the other hand the performance of the curvature mimicking portfolio was essentially as the one in Figure 3. In fact it was better as it did not include the 2008 dip. Finally, we also observed that the functional risk premium was not convex anymore, but negative and upward sloping. This is consistent with the fact that curvature may have a lesser role in this case. Similar results, though with more noise were obtained by sample splitting. In summary, the risk premium on

Table 2: Implied Variance Functional Risk Premium and Loadings. Values are only reported for a subset of moneyness. Numbers for the risk premium are multiplied by 100, hence in basis points per day.

	Moneyness %								
	80	85	90	95	100	105	110	115	120
	Risk Premium								
	-1.5	-3.1	-3.8	-3.5	-2.9	-2.7	-2.0	-1.7	-0.9
Port.	Loadings								
1	0.5	5.6	7.4	5.1	1.3	-0.5	-5.4	-7.6	-13.6
2	1.5	4.2	4.8	3.0	0.0	-1.4	-5.2	-6.9	-11.5
3	3.3	2.8	2.0	1.0	-0.6	-1.4	-3.3	-4.2	-6.7
4	2.1	2.1	1.7	0.8	-0.4	-1.0	-2.5	-3.2	-5.2
5	0.8	1.2	1.2	0.6	-0.4	-0.8	-2.0	-2.6	-4.0
6	-0.1	0.1	0.2	0.1	-0.2	-0.3	-0.6	-0.8	-1.2
7	-0.9	-0.8	-0.6	-0.4	0.0	0.1	0.5	0.7	1.3
8	-0.6	-1.4	-1.6	-1.0	-0.1	0.4	1.6	2.1	3.5
9	-1.6	-1.7	-1.5	-0.7	0.4	0.9	2.4	3.0	4.8
10	-0.7	-2.4	-2.8	-1.6	0.3	1.2	3.8	4.9	7.9
11	-2.6	-4.4	-4.5	-2.7	0.3	1.7	5.4	7.1	11.7

the curvature factor progressively decreased after the financial crisis. This observation is consistent with the fact that the performance on UMD (the momentum factor) has also deteriorated after the financial crisis.

4 Conclusion

This paper studies asset pricing relations with functional factors. Typical examples are the term structure of interest rates and the implied volatility smile. When factors are vectors, the results reduce to usual well known relations. In practice, our results naturally apply to the case of vector valued factors that are highly correlated, as it is the case for the yield curve and the options implied volatility curve. Hence, the theory allows us to extend the domain of factor based asset pricing to empirically relevant quantities in a natural way. Relying on the machinery of functional data analysis, we are able to construct sample estimators for the population quantities derived from the theory. The pricing theory suggests two elementary estimation algorithms using functional pricing factors. These algorithms are non-trivial extension of the usual two step regression procedure and the GMM estimator of risk prices using a linear discount

Table 3: Discount Factor Estimation. Results from GMM estimation minimizing the Hansen-Jagannathan Distance (11) using a diagonal scaling matrix equal to the estimated variance of the portfolios. The bootstrap 2.5% (Q2.5%) and 97.5% (Q97.5%) quantiles are also reported. These represent a 95% bootstrap confidence interval. Numbers are multiplied by 100, hence basis points per day.

	Market	Level	Slope	Curvature
b	-30.4	-145.3	0.2	1625.4
Q2.5%	-54.6	-271.2	-458.5	-916.6
Q97.5%	5.7	20.4	377.5	2669.3
λ	3.9	-2.6	0.2	0.9
Q2.5%	-0.7	-6.8	-1.0	-0.5
Q97.5%	8.3	0.9	1.0	1.7
R^2	95.9			

factor. For all the estimators, we derive consistency and central limit theorems. These results allows us to carry out inference on the sample estimates and justify the use of the bootstrap in our empirical application. All the inference results are valid for the sample size going to infinity as well as for a diverging number of assets without constraints on the rates of divergence of both.

While our results are asymptotic in the sample size, simulations results, reported in Section A.4 of the supplement, show that its applicability is justified within the context of our empirical application. Additional results can be found there.

As an empirical application we used the implied variance curve as a pricing factor. We decomposed the implied variance curve into three factor scores that can be interpreted as level, skew, and convexity. We find that curvature earns a positive risk premium.

We found that higher momentum portfolios load the curvature of the implied variance curve positively. This means that an unexpected increase in the curvature of the implied variance curve should lead to higher returns for momentum. Our results suggest that there is information in the implied variance curve and this has not been extensively exploited by past research.

References

[1] Ait-Sahalia, Y., Karaman, M., and Mancini, L. (2018) The Term Structure of Variance Swaps and Risk Premia. Journal of Econometrics 219, 204-230.

[2] Barroso, P., and Santa-Clara, P. (2015) Momentum Has Its Moments. Journal of Financial Economics 116, 111-120.

[3] Bosq, D. (2000). Linear Processes in Function Spaces: Theory and Applications, volume 149 of Lecture Notes in Statistics.

[4] Brigo, D. and Mercurio, F. (2002) Lognormal-Mixture Dynamics and Calibration to Market Volatility Smiles. International Journal of Theoretical and Applied Finance 5, 427-446.

[5] Carrasco, M., Florens, J.-P. and Renault, E. (2007) Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization. In J. Heckman and E. Leamer (eds.) Handbook of Econometrics 6B, 5633-5751. North Holland

[6] Cochrane, J.H. 2011. Presidential Address: Discount Rates. Journal of Finance 66, 1047-108.

[7] Cochrane, J.H., and Piazzesi, M. (2005) Bond Risk Premia. American Economic Review 95, 138-160.

[8] Daniel, K., and Moskowitz, T.J. (2016) Momentum Crashes. Journal of Financial Economics 122, 221-247.

[9] Della Corte, P., Kozhan, R. and Neuberger, A. (2021) The Cross-Section of Currency Volatility Premia. Journal of Financial Economics 139, 950-970.

[10] Fama, E.F., and French, K.R. (1993) Common Risk Factors in the Returns on Stocks and Bonds. Journal of Financial Economics 33, 3-56.

[11] Ferraty, F., and Vieu, P. (2006) Nonparametric Functional Data Analysis Theory and Practice Authors. Berlin: Springer.

[12] Gavish, M., and Donoho, D.L. (2014) The Optimal Hard Threshold for Singular Values is $4/\sqrt{3}$. IEEE Transactions on Information Theory 60, 5040-5053.

[13] Hong, H., Lim, T., and Stein, J.C. (2000) Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies. The Journal of Finance 55, 265-295.

[14] Horváth, L., and Kokoszka P. (2012) Inference for Functional Data With Applications. Berlin: Springer.

[15] Kargin, V. and Onatski, A. (2008) Curve Forecasting by Functional Autoregression. Journal of Multivariate Analysis 99, 2508-2526.

[16] Kokoszka, P., Miao, H. and X. Zhang (2015) Functional Dynamic Factor Model for Intraday Price Curves. Journal of Financial Econometrics 13, 456-477.

[17] Kokoszka, P., Miao, H., Reimherr, M. and Taoufik, B. (2018) Dynamic Functional Regression with Application to the Cross-Section of Returns. Journal of Financial Econometrics 16, 461-485.

[18] Kozhan, R., Neuberger, A., and Schneider, P. (2013). The skew risk premium in the equity index market. The Review of Financial Studies 26, 2174-2203.

[19] Johnstone, I.M., and Lu, A.Y. (2009) On Consistency and Sparsity for Principal Components Analysis in High Dimensions (with rejoinder). Journal of the American Statistical Association 104, 682-703.

[20] Li, D., Linton, O.B., and Lu, Z. (2015) A Flexible Semiparametric Model for Time Series. Journal of Econometrics 187, 345-357.

[21] Müller, H.G., Rituparna, S., and Stadtmüller, U. (2011) Functional Data Analysis for Volatility. Journal of Econometrics 165, 233-245.

[22] Munk, C. (2013). Financial Asset Pricing Theory. Oxford: OUP.

[23] Ramsay, J.O., and Ramsey, J.B. (2002) Functional Data Analysis of the Dynamics of the Monthly Index of Nondurable Goods Production. Journal of Econometrics 107, 327-344.

[24] Rasmussen, C. and Williams, C.K.I. (2006) Gaussian Processes of Machine Learning. Cambridge, MA: MIT Press.

[25] Sancetta (2019) Intraday End-of-Day Volume Prediction. Journal of Financial Econometrics, https://doi.org/10.1093/jjfinec/nbz019

[26] Sancetta, A. (2015) A Nonparametric Estimator for the Covariance Function of Functional Data. Econometric Theory 31, 1359-1381.

[27] Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016) Review of Functional Data Analysis. Annual Review of Statistics and Its Application 3, 257-295.

[28] Zhang, X.F. (2006) Information Uncertainty and Stock Returns. The Journal of Finance 61, 105-137.

Supplementary Material to "Empirical Asset Pricing with Functional Factors" by P. Nadler and A. Sancetta

A.1 Proof of Results

We start with the proof of Lemmas 1 and 2 and Theorem 1. We then state some preliminary lemmas that will be used to prove each of the remaining results in the paper.

A.1.1 Proof of Lemma 1

For simplicity, we drop the subscript t in the random variables. Using the spectral theorem,

$$C_{FF}(v,s) := Cov(F(v), F(s)) = \sum_{i=1}^{\infty} \theta_i \Psi_i(v) \Psi_i(s)' \tag{A.1}$$

where the functions Ψ_i are $L \times 1$ vector valued and satisfy $\langle \Psi_i, \Psi_j \rangle_{\mathcal{H}^L} = 1$ if $i = j$ and zero otherwise. The coefficients θ_i are non-negative and decreasing. To keep the notation simpler, suppose that $\mathbb{E}F = 0_L$. We can then expand

$$F(s) = \sum_{j=1}^{\infty} \sqrt{\theta_i} \zeta_i \Psi_i(s) \tag{A.2}$$

where the random variables ζ_i have mean zero and variance one and are uncorrelated across the index i. The above series expansions also hold uniformly by the multivariate version of Mercer theorem (De Vito et al., 2013). These quantities should not be confused with $\rho_i^{(l)}$, $\xi_i^{(l)}$, and $\Phi_i^{(l)}$ as defined in Theorem 1, which are derived from the scalar valued factors $F^{(l)}$ independently across $l = 1, 2, ..., L$, as opposed to F. We have that

$$Cov\left(F(s), R^{(k)}\right) = \sum_{i=1}^{\infty} \sqrt{\theta_i} \left(\mathbb{E}R^{(k)}\zeta_i\right) \Psi_i(s).$$

The definition of pricing factor says that $\beta^{(k)} \in \mathcal{H}^L$, where $\beta^{(k)}$ is the transpose of the k^{th} row of β. Since $\beta^{(k)}$ is an element in the Hilbert space $\mathcal{H}^L$, we have the representation $\beta^{(k)}(s) = \sum_{i=1}^{\infty} \tilde{\beta}_i^{(k)} \Psi_i(s)$, where the scalar coefficients $\tilde{\beta}_i^{(k)}$ satisfy $\sum_{i=1}^{\infty} \left|\tilde{\beta}_i^{(k)}\right|^2 < \infty$, uniformly in $k = 1, 2, ..., K$. From (2) and the above display, we deduce that

$$\sum_{i=1}^{\infty} \sqrt{\theta_i} \left(\mathbb{E} R^{(k)} \zeta_i \right) \gamma_i = \sum_{j=1}^{\infty} \theta_i \tilde{\beta}_i^{(k)} \gamma_i$$

for any $\gamma(s) = \sum_{i=1}^{\infty} \gamma_i \Psi_i(s)$. Given that the coefficients γ_i are arbitrary except for the fact of being square summable, this implies that $\tilde{\beta}_i^{(k)} = \theta_i^{-1/2} \left(\mathbb{E} R^{(k)} \zeta_i \right)$ for any $i \geq 1$. For $\tilde{\beta}_i^{(k)}$ to be square summable, we must have that $\sum_{i=1}^{\infty} \theta_i^{-1} \left| \mathbb{E} R^{(k)} \zeta_i \right|^2 < \infty$, uniformly in $k = 1, 2, ..., K$. Using (A.2) this summation is the same as (3). This establishes the desired implication, which clearly is in both directions.

A.1.2 Proof of Lemma 2

For simplicity we drop the subscript t in the random variables. We use the notation from Lemma 1 and its proof. From that proof recall that $\beta^{(k)}(s) = \sum_{i=1}^{\infty} \tilde{\beta}_i^{(k)} \Psi_i(s)$ where $\tilde{\beta}_i = \theta_i^{-1/2} \left(\mathbb{E} R^{(k)} \zeta_i \right)$. First, we show the only if implication, i.e. we assume that $b \in \mathcal{H}^L$. We expand $b(s) = \sum_{i=1}^{\infty} \tilde{b}_i \Psi_i(s)$ where $\sum_{i=1}^{\infty} \left| \tilde{b}_i \right|^2 < \infty$. Then,

$$\int_0^1 Cov \left(R^{(k)}, F(s) \right) b(s)\, ds = \sum_{i=1}^{\infty} \sqrt{\theta_i} \left(\mathbb{E} R^{(k)} \zeta_i \right) \tilde{b}_i.$$

Defining $\tilde{\lambda}_i := \theta_i \tilde{b}_i$, the r.h.s. can be rewritten as $\sum_{j=1}^{\infty} \tilde{\beta}_i^{(k)} \tilde{\lambda}_i$ so that $\lambda(s) = \sum_{i=1}^{\infty} \tilde{\lambda}_i \Psi_i(s)$. Dy definition of the coefficients $\tilde{\lambda}_i$, we deduce that

$$\sum_{i=1}^{\infty} \frac{\langle \lambda, \Psi_i \rangle_{\mathcal{H}^L}^2}{\theta_i^2} = \sum_{i=1}^{\infty} \left| \tilde{b}_i \right|^2 < \infty.$$

The implication on the other side requires use to use the above display, hence, it is trivial.

A.1.3 Proof of Theorem 1

For simplicity we drop the subscript t in the random variables. The existence of the second moment means that the covariance functions $C_{FF}^{(l)} = \left\{ Cov \left(F^{(l)}(s), F^{(l)}(v) \right) : s, v \in [0, 1] \right\}$,

$l = 1, 2, \ldots, L$, are well defined and continuous. By the spectral theorem,

$$C_{FF}^{(l)}(s, v) = \sum_{i=1}^{\infty} \rho_i^{(l)} \Phi_i^{(l)}(s) \Phi_i^{(l)}(v)' \tag{A.3}$$

where $\infty > \rho_i^{(l)} \geq \rho_{i+1}^{(l)} \geq 0$ for $i \geq 1$, and $\left\{\Phi_i^{(l)} : i = 1, 2, ...\right\}$ is an orthonormal basis of $\mathcal{H}$. By Mercer Theorem (Bosq, 2000), the series on the right hand side also converges uniformly. By orthonormality of the basis functions,

$$\int_0^1 C_{FF}^{(l)}(s, s)\, ds = \sum_{i=1}^{\infty} \rho_i^{(l)} = \mathbb{E}\left|F^{(l)}\right|_{\mathcal{H}}^2 < \infty \tag{A.4}$$

so that the eigenvalues are absolutely summable. This proves Point 1 of the theorem.

By the Karhunen-Loéve Theorem, for $l = 1, 2, ..., L$,

$$F^{(l)}(s) = \mathbb{E}F^{(l)}(s) + \sum_{i=1}^{\infty} \sqrt{\rho_i^{(l)}} \xi_i^{(l)} \Phi_i^{(l)}(s)$$

where $\left\{\xi_i^{(l)} \in \mathbb{R} : i = 1, 2, ...\right\}$ are mean zero, variance one, uncorrelated random variables. The second moment of the series on the right hand side converges uniformly. This proves Point 2 of the theorem.

For each $l = 1, 2, ..., L$, given that $\left\{\Phi_i^{(l)} : i = 1, 2, ...\right\}$ in (A.3) forms a basis of $\mathcal{H}$, there are scalars $\beta_i^{(k,l)}$ and $\lambda_i^{(l)}$ such that

$$\beta^{(k,l)}(s) = \sum_{i=1}^{\infty} \beta_i^{(k,l)} \Phi_i^{(l)}(s), \;\; \lambda^{(l)}(s) = \sum_{i=1}^{\infty} \lambda_i^{(l)} \Phi_i^{(l)}(s)$$

and $\sum_{i=1}^{\infty} \left(\left|\beta_i^{(k,l)}\right|^2 + \left|\lambda_i^{(l)}\right|^2\right) < \infty$, where the convergence is under the norm $|\cdot|_{\mathcal{H}}$, for $k = 1, 2, ..., K$ and $l = 1, 2, ..., L$. Then, $\int_0^1 \beta^{(k)}(s)' \lambda(s)\, ds = \sum_{l=1}^{L} \sum_{i=1}^{\infty} \beta_i^{(k,l)} \lambda_i^{(l)}$ as required. This proves Point 3 of the theorem.

By assumption $b^{(l)} \in \mathcal{H}$, hence,

$$b^{(l)}(s) = \sum_{i=1}^{\infty} b_i^{(l)} \Phi_i^{(l)}(s) \tag{A.5}$$

where the equality is under the $|\cdot|_{\mathcal{H}}$ norm, and the scalar coefficients $b_i^{(l)}$ are square summable over i for $l = 1, 2, ..., L$. By Point 2, deduce that,

$$Cov\left(R, F^{(l)}(s)\right) = \sum_{i=1}^{\infty} \sqrt{\rho_i^{(l)}} \left(\mathbb{E}R\xi_i^{(l)}\right) \Phi_i^{(l)}(s)'. \tag{A.6}$$

Using (A.6) and (A.5) in (5), deduce that

$$\mathbb{E}R = \alpha + \sum_{l=1}^{L} \sum_{i=1}^{\infty} \sqrt{\rho_i^{(l)}} \left(\mathbb{E}R\xi_i^{(l)}\right) b_i^{(l)},$$

which is the last result we needed to show.

A.1.4 Proof of Lemma 3

Here and throughout the rest of all the proofs, we shall use the notation of Lemma 3 with possibly no further mention. Lemma 3 is a consequence of Theorem 1 and the following.

Lemma 5 *Suppose that Condition 1 holds and that in the discount factor (4) only the first I factor scores are priced for each factor $l = 1, 2, ..., L$, and that their covariance matrix is full rank. Let $R_t^e = R_t - 1_K R_t^{(0)}$, where $R_t^{(0)}$ is the zero beta portfolio. Then,*

$$R_t^e = a + BS_t + \varepsilon \tag{A.7}$$

where

$$a = \left(\mathbb{E}R_t^e - B\mathbb{E}S_t\right), \; B = Cov\left(R_t, S_t\right) Var\left(S_t\right)^{-1} \tag{A.8}$$

and the error term ε_t is a mean zero vector, uncorrelated with the factor scores S_t. The equality in (A.7) holds under $\mathbb{E}\left|\cdot\right|_2^2$. If the factor scores are also tradable, $a = 0_K$.

Proof. The linear projection of $(1 - \mathbb{E}) R_t^e$ onto $(1 - \mathbb{E}) S_t$ is $B(1 - \mathbb{E}) S_t$ where $B = Cov(R_t, S_t) Var(S_t)^{-1}$. We have used the fact that, by definition of the zero beta portfolio, $Cov\left(1_K R_t^{(0)}, S_t\right)$ is zero. Hence,

$$(1 - \mathbb{E}) R_t^e = B(1 - \mathbb{E}) S_t + \varepsilon_t$$

where $\varepsilon_t = (1 - \mathbb{E}) R_t^e - B(1 - \mathbb{E}) S_t$. By construction, ε_t is mean zero and orthogonal to S_t. Rewrite the previous display as

$$R_t^e = (\mathbb{E}R_t^e - B\mathbb{E}S_t) + BS_t + \varepsilon_t.$$

The pricing relation via the discount factor implies

$$0_K = \mathbb{E}R_t^e M_t = \mathbb{E}R_t^e - \int_0^1 Cov\left(R_t, F_t\left(s\right)\right) b\left(s\right) ds = \mathbb{E}R_t^e - Cov\left(R_t, S_t\right) b_0$$

where b_0 is as in Lemma 3, because $b^{(l)}\left(s\right) = \sum_{i=1}^{I} b_i \Phi_i^{(l)}\left(s\right)$ by Theorem 1. Hence, the last two displays imply that

$$R_t = \left(BVar\left(S_t\right) b_0 - B\mathbb{E}S_t\right) + BS_t + \varepsilon_t. \tag{A.9}$$

If the factors are excess returns of tradable assets, the factors and any of their linear combinations are also priced by the discount factor, so that

$$0_{LI} = \mathbb{E}S_t M_t = \mathbb{E}S_t - Cov\left(S_t, S_t\right) b_0.$$

The above display implies that $Var\left(S_t\right) b_0 = \mathbb{E}S_t$. Inserting this in (A.9) we have that that $a = \left(BVar\left(S_t\right) b_0 - B\mathbb{E}S_t\right) = 0_K$. $\blacksquare$

A.1.5 Preliminary Lemmas for the Proof of Remaining Results

We need additional notation in order to state the results of this section. Let $\mathcal{H}^{u\times v} = \mathcal{H}^{u\times v}\left([0,1]\right)$ be the separable Hilbert space of $u \times v$ dimensional matrix valued functions on $[0,1]$ with inner product $\langle x, y\rangle_{\mathcal{H}^{u\times v}} = \int_0^1 \text{Trace}\left(x\left(s\right)' y\left(s\right)\right) ds,\ x, y \in \mathcal{H}^{u\times v}$ and norm $|\cdot|_{\mathcal{H}^{u\times v}}$ induced by the inner product. For any matrix covariance function C, its Hilbert-Schmidt norm is defined to be $|C|_{\mathcal{S}} = \sqrt{\int_0^1 \int_0^1 |C\left(r,s\right)|_F^2 \, drds}$. If C has finite Hilbert-Schmidt norm, we write $C \in \mathcal{S}$. We shall still use the same notation when C is real valued rather than matrix valued. As in the main text, recall that we use C_{RS} and similar quantities to denote the covariance between the variable in the subscripts. For any matrix A and compatible vector x, $|Ax|_2 \leq |A|_{op} |x|_2 \leq |A|_F |x|_2$. Here, $|\cdot|_{op}$ is the operator norm, i.e. the maximum singular value. Moreover, for compatible matrices A and B, $|AB|_F \leq |A|_F |B|_F$. We recall that the Frobenius norm of a matrix can be written as the sum of its squared entries. At times we shall use the basic inequality $|Cov\left(X,Y\right)|_F^2 \leq 2\left|\mathbb{E}XY'\right|_F^2$ for arbitrary, possibly vector valued random variables X

and Y.

The estimated eigenfunctions are only identified up to a sign change. With no loss of generality, we assume throughout that $\left\langle \Phi_i^{(l)}, \hat{\Phi}_i^{(l)} \right\rangle_{\mathcal{H}} > 0$ for $l = 1, 2, \ldots, L$.

Finally, recall that $\hat{\mathbb{E}}$ will denote empirical expectation over the time index t, i.e. sample average, based on a sample of size n.

We shall rely on the following simple calculation in multiple places.

Lemma 6 *Let $X := (X_t)_{t\in\mathbb{Z}}$ and $Y := (Y_t)_{t\in\mathbb{Z}}$ be $u \times 1$ and $v \times 1$ sequences of random vectors. Then, $\left|\hat{\mathbb{E}}X_t Y_t'\right|_F^2 \leq \hat{\mathbb{E}} |X_t|_2^2 \hat{\mathbb{E}} |Y_t|_2^2$ and $\left|\hat{\mathbb{E}}X_t Y_t'\right|_F^2 \leq \hat{\mathbb{E}} |X_t|_2^2 |Y_t|_2^2$*

Proof. By definition, $\left|\hat{\mathbb{E}}X_t Y_t'\right|_F^2 = \sum_{k=1}^{u} \sum_{l=1}^{v} \left|\hat{\mathbb{E}}X_{t,k} Y_{t,l}\right|^2$, where the subscript denotes the element in the vector. Applying Holder inequality, we obtain the first inequality. Conversely, using Jensen inequality we obtain the second inequality. ■

The above can easily be extended to the product of $\mathcal{H}^u$ and $\mathcal{H}^v$ valued random elements and we might do this with no further mention. Clearly, Lemma holds for expectation w.r.t. any probability measure, e.g. using $\mathbb{E}$ instead of the empirical expectation $\hat{\mathbb{E}}$.

We shall also use the L_2 law of large numbers in separable Hilbert spaces.

Lemma 7 *Suppose that $(X_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence of random variables with mean zero, finite variance, and values in a separable Hilbert space $\mathbb{H}$ equipped with the inner product $\langle\cdot,\cdot\rangle_{\mathbb{H}}$ and norm $|\cdot|_{\mathbb{H}}$, induced by this inner product. Then, $\mathbb{E}\left|\frac{1}{n}\sum_{t=1}^{n} X_t\right|_{\mathbb{H}}^2 \leq \frac{1}{n}\mathbb{E}|X_1|_{\mathbb{H}}^2$.*

Proof. This follows by the i.i.d. condition because $\mathbb{E}\langle X_s, X_t\rangle_{\mathbb{H}} = \langle \mathbb{E}X_s, \mathbb{E}X_t\rangle_{\mathbb{H}} = 0$ if $s \neq t$. ■

Note that the above result remains true if we replace the i.i.d. condition with a martingale difference condition. In this case, the l.h.s. is bounded above by $\frac{1}{n}\max_{t\leq n}\mathbb{E}|X_t|_{\mathbb{H}}^2$.

Lemma 7 shall be applied with X_t equal to F_t, $(1-\mathbb{E}) F_t F_t'$, and $(1-\mathbb{E}) A' R_t F_t'$ with the norms $|\cdot|_{\mathcal{H}^L}$, $|\cdot|_{\mathcal{S}}$, and $|\cdot|_{\mathcal{H}^{p\times L}}$, respectively, where A is as in Theorem 2. To this end, we state the following.

Lemma 8 *Under the Regularity Conditions, $\mathbb{E}|(1-\mathbb{E}) F_t F_t'|_{\mathcal{S}}^2 = O(1)$, $\mathbb{E}|(1-\mathbb{E}) A' R_t F_t'|_{\mathcal{H}^{p\times L}}^2 = O(1)$, where A is as in Theorem 2.*

Proof. It is sufficient to show that both $\mathbb{E}\left|F_tF_t'\right|_{\mathcal{S}}^2$ and $\mathbb{E}\left|A'R_tF_t'\right|_{\mathcal{H}^{p\times L}}^2$ are $O(1)$. Note that $\mathbb{E}\left|F_tF_t'\right|_{\mathcal{S}}^2=\mathbb{E}\left|F_t\right|_{\mathcal{H}^L}^4<\infty$. Moreover, $\mathbb{E}\left|A'R_tF_t'\right|_{\mathcal{H}^{p\times L}}^2=\mathbb{E}\left|A'R_t\right|_2^2\left|F_t\right|_{\mathcal{H}^L}^2$. By Holder inequality, it is sufficient to note that $\mathbb{E}\left|A'R_t\right|_2^4<\infty$ and $\mathbb{E}\left|F_t\right|_{\mathcal{H}^L}^4<\infty$ by the Regularity Conditions and the assumption on A. This concludes the proof. ■

We need the following well know result about the second order effect of using sample means when computing sample covariances. For completeness we give the proof as we are considering Hilbert valued random variables. For ease of notation, we use $\hat{\mathbb{E}}$ in what follows.

Lemma 9 *Suppose that $(X_t)_{t\in\mathbb{Z}}$ and $(Y_t)_{t\in\mathbb{Z}}$ are sequences of random variables with values in $\mathcal{H}^u$ and $\mathcal{H}^v$ respectively. Then,*

$$\hat{C}_{XY}-C_{XY}=\left(\hat{\mathbb{E}}-\mathbb{E}\right)\left[\left(1-\mathbb{E}\right)X_t\right]\left[\left(1-\mathbb{E}\right)Y_t\right]'-\left[\left(\hat{\mathbb{E}}-\mathbb{E}\right)X_t\right]\left[\left(\hat{\mathbb{E}}-\mathbb{E}\right)Y_t\right]'$$

where

$$\hat{C}_{XY}:=\hat{\mathbb{E}}\left[\left(1-\hat{\mathbb{E}}\right)X_t\right]\left[\left(1-\hat{\mathbb{E}}\right)Y_t\right]'=\hat{\mathbb{E}}\left[\left(1-\hat{\mathbb{E}}\right)X_t\right]Y_t'$$

and $C_{XY}:=Cov\left(X,Y\right)=\mathbb{E}\left[\left(1-\mathbb{E}\right)X_t\right]Y_t'$.

In particular, if the random variables are i.i.d. and $\mathbb{E}\left|X_t\right|_{\mathcal{H}^u}^2+\mathbb{E}\left|Y_t\right|_{\mathcal{H}^v}^2<\infty$, then

$$\left|\left[\left(\hat{\mathbb{E}}-\mathbb{E}\right)X_t\right]\left[\left(\hat{\mathbb{E}}-\mathbb{E}\right)Y_t\right]'\right|_{\mathcal{H}^{u\times v}}=O_p\left(\frac{1}{n}\right).$$

Proof. For the sake of clarity, we note that the equalities on the r.h.s. for the sample and population covariance, i.e. $\hat{C}_{XY}=\hat{\mathbb{E}}\left[\left(1-\hat{\mathbb{E}}\right)X_t\right]Y_t'$ and similarly for C_{XY}, are trivially verified by direct calculation. We now start the proof. Adding and subtracting $\hat{\mathbb{E}}\left[\left(1-\mathbb{E}\right)X_t\right]Y_t'$, and rearranging, we have that

$$\hat{C}_{XY}-C_{XY}=\left(\hat{\mathbb{E}}-\mathbb{E}\right)\left[\left(1-\mathbb{E}\right)X_t\right]Y_t'-\hat{\mathbb{E}}\left[\left(\hat{\mathbb{E}}-\mathbb{E}\right)X_t\right]Y_t'.$$

Now, add and subtract $\left(\hat{\mathbb{E}}-\mathbb{E}\right)\left[\left(1-\mathbb{E}\right)X_t\right]\mathbb{E}Y_t'=\left[\left(\hat{\mathbb{E}}-\mathbb{E}\right)X_t\right]\mathbb{E}Y_t'$ to find that the above display is equal to

$$\left(\hat{\mathbb{E}}-\mathbb{E}\right)\left[\left(1-\mathbb{E}\right)X_t\right]\left[\left(1-\mathbb{E}\right)Y_t\right]'-\left[\left(\hat{\mathbb{E}}-\mathbb{E}\right)X_t\right]\left[\left(\hat{\mathbb{E}}-\mathbb{E}\right)Y_t\right]'$$

and the first part of the result follows. For the second part, by definition of the norm

$|\cdot|_{\mathcal{H}^{u\times v}}$ and the Cauchy-Schwarz inequality for the Frobenius norm,

$$\left|\left[\left(\hat{\mathbb{E}}-\mathbb{E}\right)X_t\right]\left[\left(\hat{\mathbb{E}}-\mathbb{E}\right)Y_t\right]'\right|_{\mathcal{H}^{u\times v}} \leq \left|\left(\hat{\mathbb{E}}-\mathbb{E}\right)X_t\right|_{\mathcal{H}^u}\left|\left(\hat{\mathbb{E}}-\mathbb{E}\right)Y_t\right|_{\mathcal{H}^v}.$$

Then, applying Lemma 7 to each of the two terms on the r.h.s. we obtain the result. This completes the proof of the lemma. ■

The following is Corollary 4.6 in Bosq (2000).

Lemma 10 *Under Condition 2, for every $l=1,2,...,L$, $\sqrt{n}\left(\hat{C}_{FF}^{(l)}-C_{FF}^{(l)}\right)\rightarrow G_{FF}^{(l)}$ as an element in $\mathcal{S}$ where $G_{FF}^{(l)}$ is a mean zero Gaussian element with covariance operator $\Gamma_{FF}^{(l)}$ from $\mathcal{S}$ to $\mathcal{S}$ such that for any $a\in\mathcal{S}$,*

$$\left(\Gamma_{FF}^{(l)}a\right)(u,r)=Cov\left(\int_0^1\int_0^1 F^{(l)}(v)\,F^{(l)}(s)\,a(v,s)\,dvds, F^{(l)}(u)\,F^{(l)}(r)\right),\ u,r\in[0,1].$$

The following is Theorem 2.7 in Horváth and Kokoszka (2012).

Lemma 11 *Suppose that the Regularity Conditions hold. Then, $\mathbb{E}\left|\hat{\Phi}_i^{(l)}-\Phi_i^{(l)}\right|_{\mathcal{H}}^2 = O\left(n^{-1}\right)$ for $i=1,2,...,I$ and $l=1,2,...,L$.*

The following is Corollary 4.8 in Bosq (2000).

Lemma 12 *Suppose that the Regularity Conditions hold. Then, $\sqrt{n}\left(\hat{\Phi}_i^{(l)}-\Phi_i^{(l)}\right)\rightarrow G_i^{(l)}$ weakly as an element in $\mathcal{H}$, where $G_i^{(l)}$ is a mean zero Gaussian process with continuous sample paths, for $i=1,2,...,I$, $l=1,2...,L$. In particular,*

$$G_i^{(l)}(\cdot)=\int_0^1\int_0^1 D_i^{(l)}(\cdot,s)\,G_{FF}^{(l)}(s,v)\,\Phi_i^{(l)}(v)\,dsdv,$$

where $D_i^{(l)}(s,v)=\sum_{j\neq i}\left(\rho_i^{(l)}-\rho_j^{(l)}\right)^{-1}\Phi_j^{(l)}(s)\,\Phi_j^{(l)}(v)$. The coefficient $\rho_i^{(l)}$ are the eigenvalues as in (6), and $G_{FF}^{(l)}$ is as in Lemma 10.

It appears that there is a small typo in the statement of Corollary 4.8 in Bosq (2000). The expression above can also be deduced from Proposition 10 in Dauxois et al. (1982). We now show that the sample factor scores are consistent.

Lemma 13 *Suppose that the Regularity Conditions hold. Then, $\hat{\mathbb{E}}\left|\hat{S}_t-S_t\right|_2^2=O_p\left(n^{-1}\right)$.*

Proof. Note that

$$\hat{\mathbb{E}}\left|\hat{S}_t - S_t\right|_2^2 = \frac{1}{n}\sum_{t=1}^{n}\sum_{l=1}^{L}\sum_{i=1}^{I}\left|\left\langle F_t^{(l)}, \hat{\Phi}_i^{(l)} - \Phi_i^{(l)}\right\rangle_{\mathcal{H}}\right|^2.$$

By the Cauchy-Schwarz inequality, the right hand side is bounded above by

$$\sum_{l=1}^{L}\frac{1}{n}\sum_{t=1}^{n}\left|F_t^{(l)}\right|_{\mathcal{H}}^2\sum_{i=1}^{I}\left|\hat{\Phi}_i^{(l)} - \Phi_i^{(l)}\right|_{\mathcal{H}}^2.$$

By Lemma 11, $\left|\hat{\Phi}_i^{(l)} - \Phi_i^{(l)}\right|_{\mathcal{H}}^2 = O_p\left(n^{-1}\right)$. This holds for $i = 1, 2, ..., I$, $l = 1, 2, .., L$. Given that $\frac{1}{n}\sum_{t=1}^{n}\left|F_t^{(l)}\right|_{\mathcal{H}}^2 = O_p\left(1\right)$ and that L and I are bounded, we deduce the statement of the lemma. ■

We now state a series of convergence results for the sample covariance estimators.

Lemma 14 *Suppose that the Regularity Conditions hold. Then,* $K^{-1}\left|\hat{C}_{R\hat{S}} - C_{RS}\right|_F^2 = O_p\left(n^{-1}\right)$.

Proof. By linearity of the empirical covariance, adding and subtracting C_{RS}, and using $\hat{C}_{R,\left(\hat{S}-S\right)}$ to denote the sample covariance of R_t with $\hat{S}_t - S_t$,

$$K^{-1/2}\hat{C}_{R\hat{S}} = K^{-1/2}C_{RS} + K^{-1/2}\left(\hat{C}_{RS} - C_{RS}\right) + K^{-1/2}\hat{C}_{R,\left(\hat{S}-S\right)}. \tag{A.10}$$

By Lemma 9,

$$\left|K^{-1/2}\left(\hat{C}_{RS} - C_{RS}\right)\right|_F = \left|K^{-1/2}\left(\hat{\mathbb{E}} - \mathbb{E}\right)\left[\left(1-\mathbb{E}\right)R_t\right]\left[\left(1-\mathbb{E}\right)S_t'\right]\right|_F + O_P\left(n^{-1}\right)$$

if $\mathbb{E}\left|K^{-1/2}R_t\right|_2^2 + \mathbb{E}\left|S_t\right|_2^2 < \infty$. The finiteness of the two expectations follows by the Regularity Conditions because $\mathbb{E}\left|R^{(k)}\right|^4 < \infty$ and $\mathbb{E}\left(\sum_{i=1}^{\infty}\left|S_{t,i}^{(l)}\right|^2\right)^2 = \mathbb{E}\left(\sum_{i=1}^{\infty}\rho_i^{(l)}\left|\xi_{t,i}^{(l)}\right|^2\right)^2 = \mathbb{E}\left|F^{(l)}\right|_{\mathcal{H}}^4 < \infty$, $k = 1, 2, ..., K$, $l = 1, 2, ..., L$, together with the fact that L is bounded. We have actually argued that $\mathbb{E}\left|K^{-1/2}R_t\right|_2^4 + \mathbb{E}\left|S_t\right|_2^4 < \infty$, as we will need this momentarily. Using Lemma 7 again, we deduce that the first term on the r.h.s. of the previous display is $O_p\left(n^{-1/2}\right)$ if $\mathbb{E}\left|K^{-1/2}\left[\left(1-\mathbb{E}\right)R_t\right]\left[\left(1-\mathbb{E}\right)S_t'\right]\right|_F^2 < \infty$. By the Cauchy-Schwarz inequality for the Frobenius norm and Holder inequality this is the case if $\mathbb{E}\left|K^{-1/2}R_t\right|_2^4 + \mathbb{E}\left|S_t\right|_2^4 < \infty$. By the previous remarks, those expectations are finite. Hence, by the continuous mapping theorem, $\left|K^{-1/2}\left[\hat{C}_{RS} - C_{RS}\right]\right|_F^2 = O_p\left(n^{-1}\right)$.

We now bound the last term in (A.10). By definition of sample covariance, a basic inequality, and Lemma 6,

$$\frac{1}{K}\left|\hat{C}_{R,(\hat{S}-S)}\right|_F^2 \leq 2\frac{1}{K}\hat{\mathbb{E}}\left|R_t\right|_2^2\hat{\mathbb{E}}\left|\hat{S}_t - S_t\right|_2^2.$$

Using Lemma 13 and a moment bound, deduce that $K^{-1}\left|\hat{C}_{R,(\hat{S}-S)}\right|_F^2 = O_p\left(n^{-1}\right)$ because L and I are bounded. ■

Lemma 15 *Suppose that the Regularity Conditions hold. Then,* $\left|\hat{C}_{\hat{S}\hat{S}} - C_{SS}\right|_F^2 = O_p\left(n^{-1}\right)$.

Proof. By linearity, $\hat{C}_{\hat{S}\hat{S}} - C_{SS} = \hat{C}_{\hat{S},\hat{S}-S} + \hat{C}_{\hat{S}-S,S}$ where the sample covariance is computed between the objects in the subscript. By Lemma 6, and a basic inequality, we have that $\left|\hat{C}_{\hat{S},\hat{S}-S}\right|_F^2 \leq 2\hat{\mathbb{E}}\left|\hat{S}_t\right|_2^2\hat{\mathbb{E}}\left|\hat{S}_t - S_t\right|_2^2$. We note that $\hat{\mathbb{E}}\left|\hat{S}_t\right|_2^2 = O_p(1)$ because L is bounded. Hence, using Lemma 13, $\left|\hat{C}_{\hat{S},\hat{S}-S}\right|_F^2 = O_p\left(n^{-1}\right)$. The same argument applies to $\hat{C}_{S,\hat{S}-S}$. This completes the proof. ■

Lemma 16 *Suppose that the Regularity Conditions hold. Then,* $K^{-1/2}\left|\hat{B} - B\right|_F \to 0$ *in probability.*

Proof. By definition,

$$\frac{1}{K}\left|\hat{B} - B\right|_F^2 = \frac{1}{K}\left|\hat{C}_{R\hat{S}}\hat{C}_{\hat{S}\hat{S}}^{-1} - C_{RS}C_{SS}^{-1}\right|_F^2. \tag{A.11}$$

Adding and subtracting $\hat{C}_{R\hat{S}}C_{SS}^{-1}$, and using a standard inequality, the above is bounded above by two times

$$\frac{1}{K}\left|\hat{C}_{R\hat{S}}\left(\hat{C}_{\hat{S}\hat{S}}^{-1} - C_{SS}^{-1}\right)\right|_F^2 + \frac{1}{K}\left|\left(\hat{C}_{R\hat{S}} - C_{RS}\right)C_{SS}^{-1}\right|_F^2. \tag{A.12}$$

By Lemma 14, we can replace $K^{-1/2}\hat{C}_{R\hat{S}}$ with $K^{-1/2}C_{RS}$ on an event that has probability going to one. By the Regularity Conditions, the latter has largest eigenvalue bounded away from infinity. Then, by Lemma 15 and the fact that C_{SS} is full rank by Condition 5, we can replace $\hat{C}_{\hat{S}\hat{S}}^{-1}$ with $\hat{C}_{SS}^{-1}$ so that the first term in the above display goes to zero in probability. The second term also goes to zero in probability by Lemma 14 because C_{SS} has smallest eigenvalue bounded away from zero using the Regularity Conditions. ■

Lemma 17 *Suppose that the Regularity Conditions hold. For A as in Theorem 2, we have that $\sqrt{n}A'\hat{C}_{R,\hat{S}-S} \to G_{A,1}$ in distribution, where $G_{A,1}$ is a mean zero random matrix and the convergence also holds for $K \to \infty$.*

Proof. Note that $\sqrt{n}A\hat{C}_{R,\hat{S}-S} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} A'\left[\left(1-\hat{\mathbb{E}}\right)R_t\right]\left(\hat{S}_t - S_t\right)$ so that the $i+(l-1)I$ column of the r.h.s. is equal to

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} A'\left[\left(1-\hat{\mathbb{E}}\right)R_t\right]\left\langle F_t^{(l)}, \hat{\Phi}_i^{(l)} - \Phi_i^{(l)}\right\rangle_{\mathcal{H}}.$$

This display is a $p\times 1$ vector and is equal to

$$\int_0^1 \frac{1}{n}\sum_{t=1}^{n} A'\left[\left(1-\hat{\mathbb{E}}\right)R_t\right] F_t^{(l)}(s)\sqrt{n}\left(\hat{\Phi}_i^{(l)}(s) - \Phi_i^{(l)}(s)\right)ds \tag{A.13}$$

By Lemma 12, $\sqrt{n}\left(\hat{\Phi}_i^{(l)} - \Phi_i^{(l)}\right)$ converges weakly to a mean zero Gaussian process $G_i^{(l)}$ in $\mathcal{H}$, and the convergence is joint in $l = 1,2,\ldots,L$ and $i = 1,2,..,I$ as long as LI is bounded. Then the proof is complete if we can show that each element in the sample average in (A.13) is in $\mathcal{H}$. Write

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n} A'\left[\left(1-\hat{\mathbb{E}}\right)R_t\right]F_t(s)'\\ =&\frac{1}{n}\sum_{t=1}^{n} A'\left[\left(\mathbb{E}-\hat{\mathbb{E}}\right)R_t\right]F_t(s)' + \frac{1}{n}\sum_{t=1}^{n} A'\left[\left(1-\mathbb{E}\right)R_t\right]F_t(s)'.\end{aligned} \tag{A.14}$$

The equality follows adding and subtracting $A'(\mathbb{E}R_t)F_t(s)'$. It is sufficient to show that each of the terms on the r.h.s. is in $\mathcal{H}^{p\times L}$. Using a simple extension of Lemma 6, we have that

$$\left|\frac{1}{n}\sum_{t=1}^{n} A'\left[\left(\mathbb{E}-\hat{\mathbb{E}}\right)R_t\right]F_t'\right|_{\mathcal{H}^{p\times L}} \leq \hat{\mathbb{E}}\left|\left(\mathbb{E}-\hat{\mathbb{E}}\right)A'R_t\right|_2^2 \hat{\mathbb{E}}\left|F_t\right|_{\mathcal{H}^L}.$$

We note that $\mathbb{E}\left|\left(\mathbb{E}-\hat{\mathbb{E}}\right)A'R_t\right|_2^2 \leq n^{-1}\mathbb{E}\left|A'R_t\right|_2^2 = O\left(n^{-1}\right)$ by Lemma 7 and the assumption on A. Hence, the last display is $O_p\left(n^{-1}\right)$. This implies that the first term on the r.h.s. of (A.14) is asymptotically zero in $\mathcal{H}^{p\times L}$, in probability. Now, note that

$\mathbb{E}A' [(1-\mathbb{E}) R_t] F_t(s)' = A'C_{RF}$. By Lemmas 7 and 8,

$$\mathbb{E}\left|\frac{1}{n}\sum_{t=1}^{n}\left[\left(\left(1-\mathbb{E}\right)A'R_t\right)F_t' - A'C_{RF}\right]\right|_{\mathcal{H}^{p\times L}} \to 0.$$

These calculations imply that the l.h.s. of (A.14) converges to $A'C_{RF}$ as an element in $\mathcal{H}^{p\times L}$. This implies that each entry in $A'\hat{C}_{RF}$ is in $\mathcal{H}$ with probability going to one, and this completes the proof. ■

Lemma 18 *Suppose that the Regularity Conditions hold. For A as in Theorem 2, we have that $\sqrt{n}A'\left(\hat{C}_{RS} - C_{RS}\right) \to G_{A,2}$ in distribution, where $G_{A,2}$ is a mean zero random matrix and the convergence also holds for $K \to \infty$.*

Proof. By Lemma 9, $\sqrt{n}A'\left(\hat{C}_{RS} - C_{RS}\right)$ has same distribution as

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\left(1-\mathbb{E}\right)A'R_t\right]\left[\left(1-\mathbb{E}\right)S_t\right]'$$

if $\mathbb{E}\left|A'R_t\right|_2^2 + \mathbb{E}\left|S_t\right|_2^2 < \infty$. As shown in the proof of Lemma 14, this is the case. Hence, by the i.i.d. condition, for the central limit theorem to apply to the above display, it is sufficient to show that $\mathbb{E}\left|A'R_tS_t'\right|_F^2 < \infty$ (Aldous, 1976). By Holder inequality, this moment bound is implied by $\mathbb{E}\left|S_t\right|_2^4 \leq \mathbb{E}\left|F_t\right|_{\mathcal{H}^L}^4 < \infty$ and $\mathbb{E}\left|A'R_t\right|_2^4 < \infty$. The finiteness of the latter quantities is shown in the proof of Lemma 14. When $K \to \infty$ we actually need uniform integrability of $\mathbb{E}\left|A'R_t\right|_2^4$, but this holds by the conditions of Theorem 2. Hence the above display converges to a Gaussian random matrix. ■

We need to control the eigenvalues of $C_{RS}'W^jC_{RS}$ for $j = 1, 2$; here $W^j = WW^{j-1}$. To this end, we state the following standard results (Bathia, 1997, III.20 p.72).

Lemma 19 *Let A and B be two compatible matrices with singular values in $[\underline{a}, \overline{a}]$ and $\left[\underline{b}, \overline{b}\right]$, respectively. Then, the singular values of AB are bounded above and below by $\overline{a}\,\overline{b}$ and $\underline{a}\,\underline{b}$, respectively.*

Using Lemma 19, we deduce the following.

Lemma 20 *Suppose that Conditions 5 and 6 hold. Then, for any finite integer j, $C_{RS}'W^jC_{RS}$ has minimum and maximum eigenvalues bounded below and above by a constant multiple of K.*

Proof. Because of Lemma 19 and the fact that $K^{-1/2}C_{RS}$ has singular values bounded away from zero and infinity, it is sufficient to show that the eigenvalues of W are bounded away from zero and infinity. This follows by Condition 6 because a diagonal matrix has eigenvalues equal to its diagonal. Note that the eigenvalues will grow (shrink) as a power of j. As long as j is finite, the eigenvalues will remain bounded away from zero and infinity. ■

A.1.6 Proof of Theorem 2

Using a decomposition analogous to (A.10), it is easy to see that the result follows from Lemmas 17 and 18 together with Lemma 15, the continuous mapping theorem and Slutsky theorem.

A.1.7 Proof of Theorem 3

It is sufficient to show convergence in distribution as the dimension of $\hat{\Lambda}$ is fixed and finite. By Lemma 3, $\mathbb{E}R_t^e = B\Lambda$ so that

$$\hat{\mathbb{E}}R_t^e = \hat{B}\Lambda + \left(B - \hat{B}\right)\Lambda + \left(\hat{\mathbb{E}} - \mathbb{E}\right)R_t^e.$$

By the definition of $\hat{\Lambda}$, adding and subtracting Λ, we have that

$$\hat{\Lambda} = \Lambda + \left(\frac{\hat{B}'\hat{B}}{K}\right)^{-1}\frac{\hat{B}'}{K}\left[\left(B - \hat{B}\right)\Lambda + \left(\hat{\mathbb{E}} - \mathbb{E}\right)R_t^e\right].$$

By Lemma 16, we can replace $\left(\hat{B}'\hat{B}/K\right)^{-1}K^{-1}\hat{B}'$ with $\left(B'B/K\right)^{-1}K^{-1}B'$ using the continuous mapping theorem because $\frac{B'B}{K}$ has minimal eigenvalue bounded away from zero. The latter remark follows from the condition on the singular values of $C_{RS}/K^{1/2}$, the invertibility of C_{SS}, and (A.8) in Lemma 5. Then, subtracting Λ on both sides and multiplying by $\sqrt{n}$ we deduce that

$$\sqrt{n}\left(\hat{\Lambda} - \Lambda\right) = \left(\frac{B'B}{K}\right)^{-1}\frac{B'}{K}\left[\sqrt{n}\left(B - \hat{B}\right)\Lambda + \sqrt{n}\left(\hat{\mathbb{E}} - \mathbb{E}\right)R_t^e\right] + o_p(1). \quad \text{(A.15)}$$

We define $A := \frac{B}{K}\left(\frac{B'B}{K}\right)^{-1}$ and show that such A satisfies the condition of Theorem 2. To this end, it is sufficient to check that $\mathbb{E}\left|A'R_t\right|_2^4 < \infty$. Now, $K^{1/2}A = \frac{B}{K^{1/2}}\left(\frac{B'B}{K}\right)^{-1}$

has largest singular value, say ρ_B, bounded away from infinity. Hence,

$$\mathbb{E}\left(K^{-1}R_t'\left(KAA'\right)R_t\right)^2 = \mathbb{E}\left(\rho_B^2 K^{-1}R_t'R_t\right)^2.$$

It is easy to see that the above display is finite as long as $\max_{k\leq K}\mathbb{E}\left|R_t^{(k)}\right|^4 < \infty$, which is the case by the Regularity Conditions. Hence, A satisfies the conditions of Theorem 2 and the first term in (A.15) converges in distribution to a Gaussian vector by Theorem 2. The second term in (A.15) is $\sqrt{n}\left(\hat{\mathbb{E}} - \mathbb{E}\right)A'R_t^e$ and its convergence to a $p \times 1$ Gaussian random vector follows by the same argument as in the proof of Lemma 18. In consequence we deduce the result.

A.1.8 Proof of Theorem 4

To ease the notation, write R_t for R_t^e throughout. The solution to (11), when only the first I factor scores are priced for each factor, is the standard generalised least square estimator

$$\hat{b}_0 = \left[\hat{C}_{R\hat{S}}'\hat{W}\hat{C}_{R\hat{S}}\right]^{-1}\hat{C}_{R\hat{S}}'\hat{W}\hat{\mathbb{E}}R_t, \tag{A.16}$$

which is $LI \times 1$. From (A.16), we have that

$$\left(\hat{b}_0 - b_0\right) = \left[\frac{\hat{C}_{R\hat{S}}'\hat{W}\hat{C}_{R\hat{S}}}{K}\right]^{-1}\frac{\hat{C}_{R\hat{S}}'\hat{W}\left(\hat{\mathbb{E}}R_t - \hat{C}_{R\hat{S}}b_0\right)}{K}.$$

By Lemma 14, we can replace $\hat{C}_{R\hat{S}}$ with $\hat{C}_{RS}$ throughout. Moreover, $\hat{W}$ is diagonal and consistent for W under the Frobenius norm. Hence, by the continuous mapping theorem, the $LI \times K$ matrix

$$\left[\frac{\hat{C}_{R\hat{S}}'\hat{W}\hat{C}_{R\hat{S}}}{K}\right]^{-1}\frac{\hat{C}_{R\hat{S}}'\hat{W}}{K}$$

can be replaced with the matrix

$$\left[\frac{C_{RS}'WC_{RS}}{K}\right]^{-1}\frac{C_{RS}'W}{K}.$$

Now, by Lemma 20, the above display has singular values proportional to $K^{-1/2}$. To ease notation, define $Q_K = C_{RS}'W/K$, which is an $LI \times K$ matrix. By Lemma 20, this matrix has singular values bounded by a constant multiple of $K^{-1/2}$. Hence, we claim that that Q_K satisfies the conditions of A as in Theorem 2. To see this, we note that LI is a fixed finite number and that $\mathbb{E}\left|Q_K R_t\right|_2^4 = \mathbb{E}\left|R_t' Q_K' Q_K R_t\right|^2 = O\left(\mathbb{E}\left(K^{-1} R_t' R_t\right)^2\right)$ by the aforementioned remarks on the singular values of Q_K. Given that $\mathbb{E}\left(K^{-1} R_t' R_t\right)^2 < \infty$, we have proved the claim. Then, we shall now focus on finding the asymptotic distribution of

$$\begin{aligned}\sqrt{n} Q_K \left(\hat{\mathbb{E}} R_t - \hat{C}_{R\hat{S}} b_0\right) &= \sqrt{n}\left(\hat{\mathbb{E}} Q_K R_t - Q_K C_{RS} b_0\right) + \sqrt{n} Q_K \left(\hat{C}_{RS} - C_{RS}\right) b_0 \\ &\quad + \sqrt{n} Q_K \hat{C}_{R\left(\hat{S}-S\right)} b_0. \end{aligned} \tag{A.17}$$

The first term on the r.h.s. of (A.17) has mean zero because $\mathbb{E} R_t = C_{RS} b_0$ by Lemma 3. Hence,

$$\sqrt{n}\left(\hat{\mathbb{E}} Q_K R_t - Q_K C_{RS} b_0\right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} Q_K \left(1 - \mathbb{E}\right) R_t.$$

By the same argument as in the proof of Lemma 18, the central limit theorem applies to the above display. By Lemma 18, also the second term in (A.17) satisfies the central limit theorem. Finally, by Lemma 15, the third term in (A.17) also converges in distribution to a Gaussian vector. This completes the proof.

A.1.9 Proof of Lemma 4

Write

$$\begin{aligned}\sqrt{n}\left(\sum_{i=1}^{I} \hat{a}_i^{(l)} \hat{\Phi}_i^{(l)} - \sum_{i=1}^{I} a_i^{(l)} \Phi_i^{(l)}\right) = & \sum_{i=1}^{I} \sqrt{n}\left(\hat{a}_i^{(l)} - a_i^{(l)}\right) \hat{\Phi}_i^{(l)} \\ & + \sum_{i=1}^{I} a_i^{(l)} \sqrt{n}\left(\hat{\Phi}_i^{(l)} - \Phi_i^{(l)}\right).\end{aligned}$$

By the assumed convergence in distribution of the coefficients and Lemma 11, the first term on the r.h.s. converges to a mean zero Gaussian process. By Lemma 12 also the second term converges to a Gaussian process with mean zero. By Slutsky theorem, the result follows.

A.2 Bootstrap Confidence Intervals and Standard Errors

Our estimations are based on intermediate results, as we estimate the eigenfunctions and compute the estimated factor scores. These intermediate estimations do not alter the asymptotic distribution. However, they result in standard errors that are difficult to compute. For this reason we use the following bootstrap procedure.

1. Given the sample $\{(R_t, F_t) : t = 1, 2, ..., n\}$ resample n observations with replacement.

2. Compute the factor scores and carry out all the estimations to derive a value for loadings, risk prices and risk premia.

3. Repeat 1000 times

4. Center all the 1000 bootstrap estimates using the estimates from the original sample.

5. Compute the standard deviation assuming zero mean and divide by $\sqrt{n}$ to find the bootstrap standard errors. Alternatively, compute confidence intervals from the empirical distribution of the bootstrap distribution.

The bootstrap is valid if the asymptotic distribution of the statistic is normal and if the test statistic is smooth. The validity of this procedure is guaranteed by the convergence result in Section 2.7.2.

A.3 Additional Details on Empirical Results

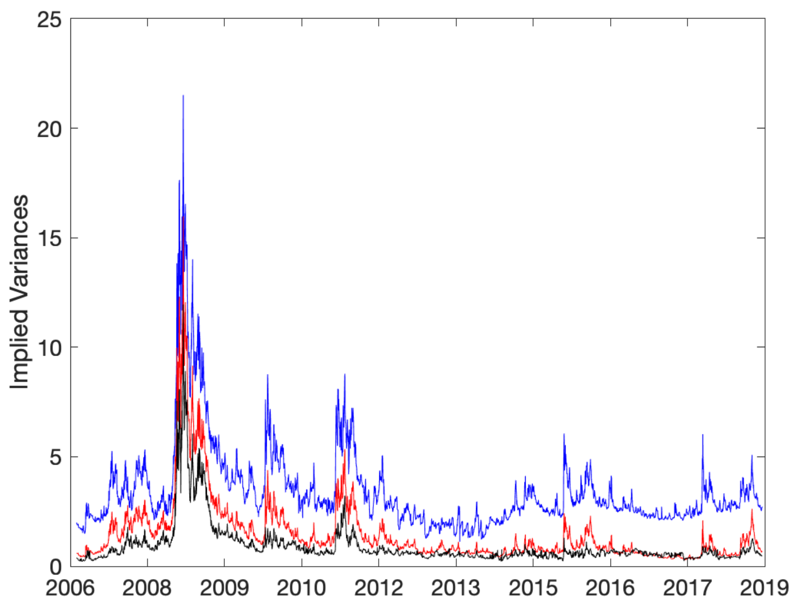
A.3.1 The Implied Vol Data

The implied volatilities were obtained from Bloomberg using the ticker: "SPX 3M m VOL LIVE INDEX". Here, m is the corresponding moneyness value from 80 to 120 per cent. The implied volatility surface is calculated by Bloomberg's Listed Implied Volatility Engine (LIVE). After estimating the corresponding implied forwards LIVE calculates a grid of implied volatility points for the option chain of each listed expiration date. The underlying model used for such calibrations is the lognormal mixture model of Brigo and Mercurio (2002). Implied volatility is quoted annualized. Since we conduct our analysis on a daily frequency we convert the implied volatility to daily volatility by dividing the original data by $\sqrt{252}$.

A.3.2 Implied Variance Statistics

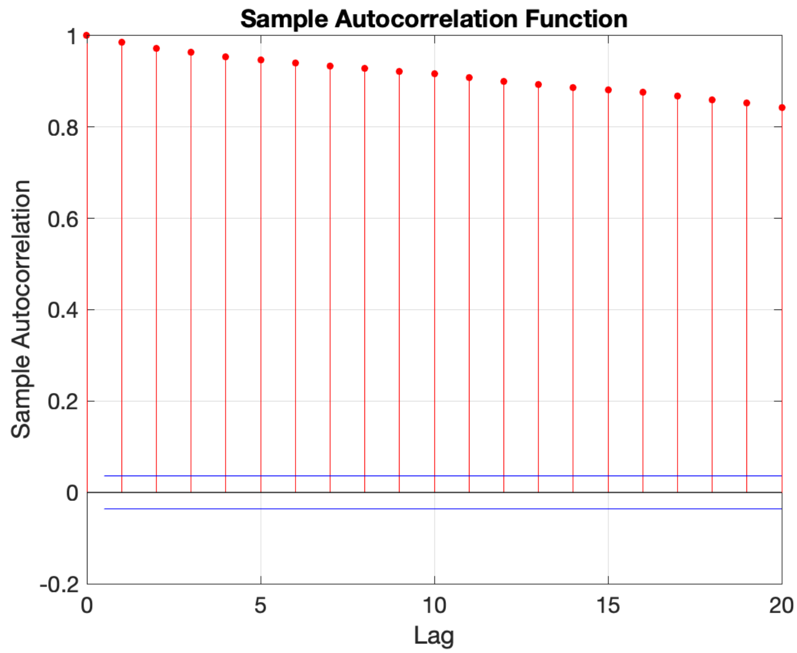
We plot the time series of the implied variance at different levels of moneyness and find it to be persistent (Figure A.1).

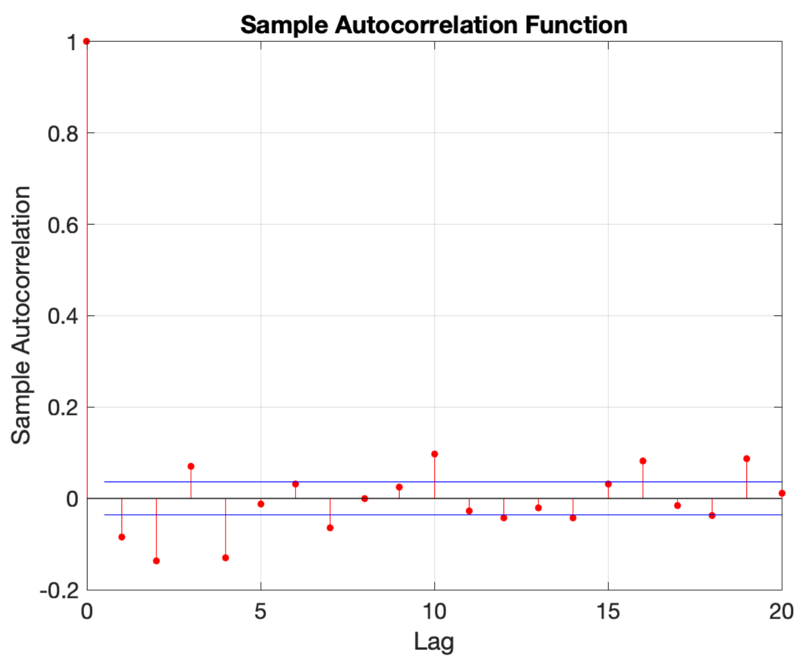
Figure A.1: Implied Variances from end of 2006 to 2019. The 80% (top line), the 100% (middle line) and the 120% (bottom line) implied variances are plotted against time at daily frequency.



The plots for the autocorrelation function of the at the money variance and variance changes suggest time series dynamics that can be approximated by a local random walk (Figure A.2).

Figure A.2: ACF. Autocorrelation of implied variance in levels (top panel) and first differences (bottom panel).





A.3.3 Descriptive Statistics of Returns and Factor Scores

Table A.1 reports yearly averages of the main quantities of interest. The results are difficult to interpret. Hence, we also compute correlations in Table A.2. We find that the market is always negatively correlated with the level of variance. During the 2008-2009 crisis period, the slope was positively correlated with the market. However, in all subsequent years, the correlation was negative. The opposite appears to be true for the curvature. This seems to suggest that the skew has essentially decreased during the years, while the curvature has increased. Given that in a year, there are about 252 trading days, the standard error for a zero correlation is approximately 0.063 (numbers in Table A.2 are multiplied by 100). In summary, we conclude that higher values of the market and curvature are associated to good states of the world, while the reverse is for level and slope.

Table A.1: Yearly Means. Average daily values of excess market returns (Mrk), the level factor score (Lvl), the slope factor score (Slp) and curvature factor score (Crv) of implied variance, and UMD. The overall mean (Overall) using the full sample of daily observations is also reported for each variable. Given that for 2019 we only have 2 months, row "Overall" does not coincide with the average over the years. Numbers are multiplied by 100 and represent basis points per day.

Year	Mrk	Lvl	Slp	Crv	UMD
2007	0.91	0.51	0.01	-0.09	9.39
2008	-15.48	1.48	-0.01	0.00	6.44
2009	11.31	-1.57	0.02	0.05	-26.24
2010	7.00	-0.14	-0.01	0.03	2.58
2011	1.34	0.31	0.07	-0.06	4.53
2012	6.36	-0.39	-0.05	0.04	-0.09
2013	12.22	-0.20	-0.07	0.01	2.83
2014	4.68	0.26	0.15	0.05	1.31
2015	0.51	-0.04	-0.01	-0.03	7.12
2016	5.30	-0.10	-0.07	0.03	-7.71
2017	7.88	-0.08	0.06	0.05	3.39
2018	-2.20	0.46	-0.04	-0.12	5.14
2019	28.95	-2.53	0.06	0.37	-20.10
Overall	3.63	0.01	0.00	0.00	0.45

We compute the contribution of the first factor score of the variance innovation curve. We do so over time, as we go from market stress into a bull market. To this end,

Table A.2: Yearly Correlation. Correlations are computed for excess market returns (Mrk) and the level factor score (Lvl) with the slope factor score (Slp) and curvature factor score (Crv) of implied variance, and the momentum factor (UMD). The overall correlation (Overall) using the full sample of daily observations is also reported for each pair of variables. Numbers are multiplied by 100 and rounded the the nearest integer.

Year	Mrk,Lvl	Mrk,Slp	Mrk,Crv	Mrk,UMD	Lvl,Slp	Lvl,Crv	Lvl,UMD
2007	-87	-28	55	5	41	-40	-13
2008	-85	45	-20	-63	-48	34	48
2009	-81	30	17	-77	-39	-12	68
2010	-81	-43	31	68	65	-30	-61
2011	-81	-31	33	-13	47	-22	7
2012	-78	-48	36	-52	70	-19	40
2013	-72	-41	29	47	71	13	-38
2014	-79	-44	34	46	66	-10	-28
2015	-88	-63	47	-9	76	-30	0
2016	-89	-73	72	-44	82	-64	32
2017	-76	-66	22	37	74	3	-35
2018	-88	-73	56	39	86	-53	-36
2019	-87	-23	70	-24	37	-77	34
Overall	-80	-11	20	-35	0	0	33

Table A.3: First Eigenvalue Contribution. The ratio of the first eigenvalue of the implied variance curve over the total sum of the eigenvalues is computed for each year. Numbers are multiplied by 100 and rounded the nearest integer.

Year	'07	'08	'09	'10	'11	'12	'13	'14	'15	'16	'17	'18	'19
Ratio	94	99	99	97	96	89	82	84	94	94	82	93	91

we report the ratio between the first eigenvalue and the total sum of the eigenvalues. This calculation is repeated for each year. We expect the level to dominate in periods of market stress consistently with our previous remarks. Table A.3 shows that this is the case. Hence, any unexpected change in the curve during the crises was mostly captured by the level.

A.3.4 The Cross-Sectional Regression

In Table A.4, taking into account the standard errors, we see that the loadings for the constant and the level are statistically small. The market loading dominates, but as expected, UMD (portfolio 11) has the smallest loading on the market. The loadings on slope and curvature provide interesting insights. The higher the momentum, the more negative is the loading on slope with reverse sign for lower momentum portfolios, resembling a decreasing linear function. This means that an unexpected increase in slope (higher out of the money put variance) reduces the return on momentum strategies. An increase in the slope means an increase in the implied variance of out of the money puts relative to the out of the money calls (i.e. a more negative slope). Conversely, the loading on curvature is positive. We can see that UMD (portfolio 11) is the one with the highest loading. UMD is close to being market neutral but still benefits from this increase in out of the money calls. We view this increase in curvature as "good variance" as opposed to the variance of the level factor which is associated with market distress.

Table A.4: Time Series Regression. The excess returns of the 11 portfolios are regressed on an intercept (const), the market excess return (Mrk) and the first three factor scores (Lvl, Slp, Crv) of the implied variance innovations. Standard errors (s.e.) are computed using the bootstrap. Numbers are multiplied by 100 and rounded to the nearest integer.

	Portfolio										
	1	2	3	4	5	6	7	8	9	10	11
const	-4	-1	0	0	1	0	1	1	0	1	1
s.e.	3	2	1	1	1	1	1	1	1	1	2
Mrk	175	138	118	110	102	97	93	92	98	109	-28
s.e.	8	4	3	2	2	1	1	1	2	3	4
Lvl	61	-14	-12	-12	-29	-21	-14	-2	19	43	18
s.e.	33	20	16	11	8	6	6	7	8	12	18
Slp	494	443	308	230	162	42	-62	-138	-205	-300	-474
s.e.	94	57	47	28	24	20	18	22	24	37	55
Crv	-439	-306	-112	-97	-86	-28	24	95	92	199	282
s.e.	121	84	55	39	34	30	36	41	41	70	79
R^2	67	80	86	90	93	95	95	92	89	76	18

A.3.5 Comparison to Carhart Four Factor Model

Carhart four-factor model (Carhart, 1997) is used here as a benchmark to put the results of the paper into perspective. Once we retain the factor scores from the implied variance surface, the number of factors is four as in Carhart four-factor model. The models are non-nested. The results from the factor loadings estimation are reported in Table A.5. We use the estimates to compute the regression in the cross-section. For the sample in question, Carhart four-factor model does not price the cross section better than our implied variance model. This can be seen comparing the R^2 in Tables A.6 and A.7 with the ones in Tables 1 and 3. It is worth noting that we were able to produce similar results to Carhart four-factor model without the use of UMD as factor. UMD has been specifically designed to capture the risk premium on momentum.

Table A.5: Time Series Regression. The excess returns on the 11 portfolios are regressed on an intercept (α), the market excess return (Mrk) and small minus big (SMB), high minus low (HML) and UMD. Standard errors (s.e.) are computed using the bootstrap. Numbers are multiplied by 100 and rounded to the nearest integer.

	Portfolio										
	1	2	3	4	5	6	7	8	9	10	11
α	0	1	1	0	1	1	-1	0	0	0	0
s.e.	1	1	1	1	0	0	0	1	0	1	0
Mrk	101	100	100	99	98	100	105	119	0	119	0
s.e.	1	1	1	1	1	1	1	2	0	2	0
SMB	5	-1	-6	-3	-3	2	5	24	0	24	0
s.e.	1	2	2	2	2	1	1	3	0	3	0
HML	1	5	5	2	1	5	5	-1	0	-1	0
s.e.	2	2	2	2	1	2	1	3	0	3	0
UMD	-58	-34	-22	-5	9	25	37	60	100	60	100
s.e.	1	1	1	1	1	1	1	2	0	2	0
R^2	97	95	95	95	95	95	95	90	100	90	100

Table A.6: Cross-Section Regression. The average return of the 11 portfolios are regressed on a pricing error and the estimated betas from the time series regression. The bootstrap 2.5% (Q2.5%) and 97.5% (Q97.5%) quantiles are also reported. These represent a 95% bootstrap confidence interval. Numbers are multiplied by 100, hence in basis points per day.

	Pricing Error	Market	SMB	HML	UMD
Risk Premia	0.1	4.1	-2.6	-5.0	0.0
Q2.5%	-0.8	-0.3	-10.1	-16.2	-3.7
Q97.5%	1.0	8.3	5.8	5.7	3.7
R^2	91.9				

Table A.7: Discount Factor Estimation. Results from GMM estimation minimizing the Hansen-Jagannathan Distance (11) using a diagonal scaling matrix equal to the estimated variance of the portfolios. The bootstrap 2.5% (Q2.5%) and 97.5% (Q97.5%) quantiles are also reported. These represent a 95% bootstrap confidence interval. Numbers are multiplied by 100, hence in basis points per day.

	Market	SMB	HML	UMD
b	5.2	-10	-14.9	-3.4
Q2.5%	0.4	-33.4	-41.8	-13.7
Q97.5%	10	16.9	9.5	6.9
λ	4.2	-2.8	-4.4	0.1
Q2.5%	0	-10.6	-13.2	-3.7
Q97.5%	8.3	5.4	4.4	3.6
R^2	92.7			

A.3.6 Additional Details on Factor Mimicking Portfolios

The pricing equation (1) still holds if we use factor mimicking portfolios in place of the non tradable factors (e.g. Breeden, 1979, Breeden et al., 1989, Huberman et al., 1987). When the factor is a curve, in principle, we may need an infinite number of portfolios for the argument to be valid. However, due to the fact that the covariance of the functional data is the kernel of a compact operator (it is square integrable), we can in principle use a small number of portfolios as approximation. This is clearly the case if only a small number of factor scores is priced. Then, we only need to mimic these factor scores. Let S_t be the three dimensional vector of factor scores for $\{\Sigma_t(s) : s \in [0,1]\}$. Then, we use the 11 portfolios to construct a factor mimicking portfolio for S_t. In particular,

$$S_t = \gamma_0 + \gamma_1' R_t + \varepsilon_t \tag{A.18}$$

where γ_0 and γ_1 are 3×1 and $K\times 3$ dimensional matrices, and the error term ε_t is a 3×1 mean zero vector when conditioning on R_t. Table A.8 reports estimation details for the constant, and portfolios 1, 10 and 11 (UMD). We note that the level loads positively portfolio 1, which has the lowest momentum. The bootstrap confidence intervals suggest little relation to portfolios 10 and UMD. The slope is the one that is most related to UMD. Given the overall poor performance of UMD over the sample, we understand why the slope does not gain a risk premium. On the other hand, as expected, the curvature does load significantly the highest momentum portfolio. Results are in Table A.8.

We construct tradable portfolios standardizing the γ_1 matrix to have columns with unit Euclidean norm. Given that the i^{th} column in γ_1 is the exposure to the portfolios returns for the i^{th} factor score, this ensure a homogeneous exposure across factor scores.[12]

Table A.8: Factor Mimicking Portfolios. Results from the regression in (A.18). The regression was carried out using R_t (11 portfolios), but only the values for the intercept and the estimated entry in γ_1 corresponding to the coefficient of UMD ($\gamma_{1,umd}$) are reported. The bootstrap 2.5% (Q2.5%) and 97.5% (Q97.5%) quantiles are also reported. These represent a 95% bootstrap confidence interval. Numbers are multiplied by 100.

	Level	Slope	Curvature
γ_0	0.82	0.02	-0.02
Q2.5%	0.23	-0.16	-0.11
Q97.5%	1.37	0.18	0.07
$\gamma_{1,port1}$	2.04	-0.19	0.00
Q2.5%	0.05	-0.58	-0.12
Q97.5%	3.70	0.24	0.13
-	-	-	-
$\gamma_{1,port10}$	-0.27	-0.29	0.25
Q2.5%	-2.62	-0.79	0.05
Q97.5%	2.06	0.22	0.46
$\gamma_{1,port11}$	-0.45	1.04	-0.19
Q2.5%	-7.28	-0.45	-0.66
Q97.5%	5.99	2.58	0.29
R^2	65.55	9.36	5.64

Table A.9, reports summary statistics for the excess returns on the market, UMD, the highest decile portfolio (portfolio 10) and the scaled curvature mimicking portfolio. UMD has not fared well during the sample period. The market, portfolio 10, and the curve mimicking portfolios enjoyed higher levels of returns. The low correlation between portfolio 10 and UMD suggest that the performance of portfolio 10 was not driven by UMD. Both portfolio 10 and the curvature mimicking portfolios are highly correlated with the market. The market remains the main pricing factor for portfolio 10. However, the correlation between portfolio 10 and the curvature mimicking portfolio

[12] This means that the portfolio weights do not add to one, as it is usually the case. However, this has no consequences for empirical results, as the portfolio remains factor mimicking (Huberman et al., 1987, Proposition 1).

Table A.9: Summary Statistics of Scaled Curvature Mimicking Portfolio Excess Returns and Other Portfolios. Mean, standard deviation (std) and Sharpe ratio are in annualized terms for the market (Mrk), UMD, the highest decile portfolio (Port10) and the scaled curvature mimicking portfolio (Crv). Here, Crv is scaled by the Euclidean norm of the portfolios γ_1 loadings. The correlations (corr) between these portfolios are also reported with numbers multiplied by 100 and rounded to the nearest integer.

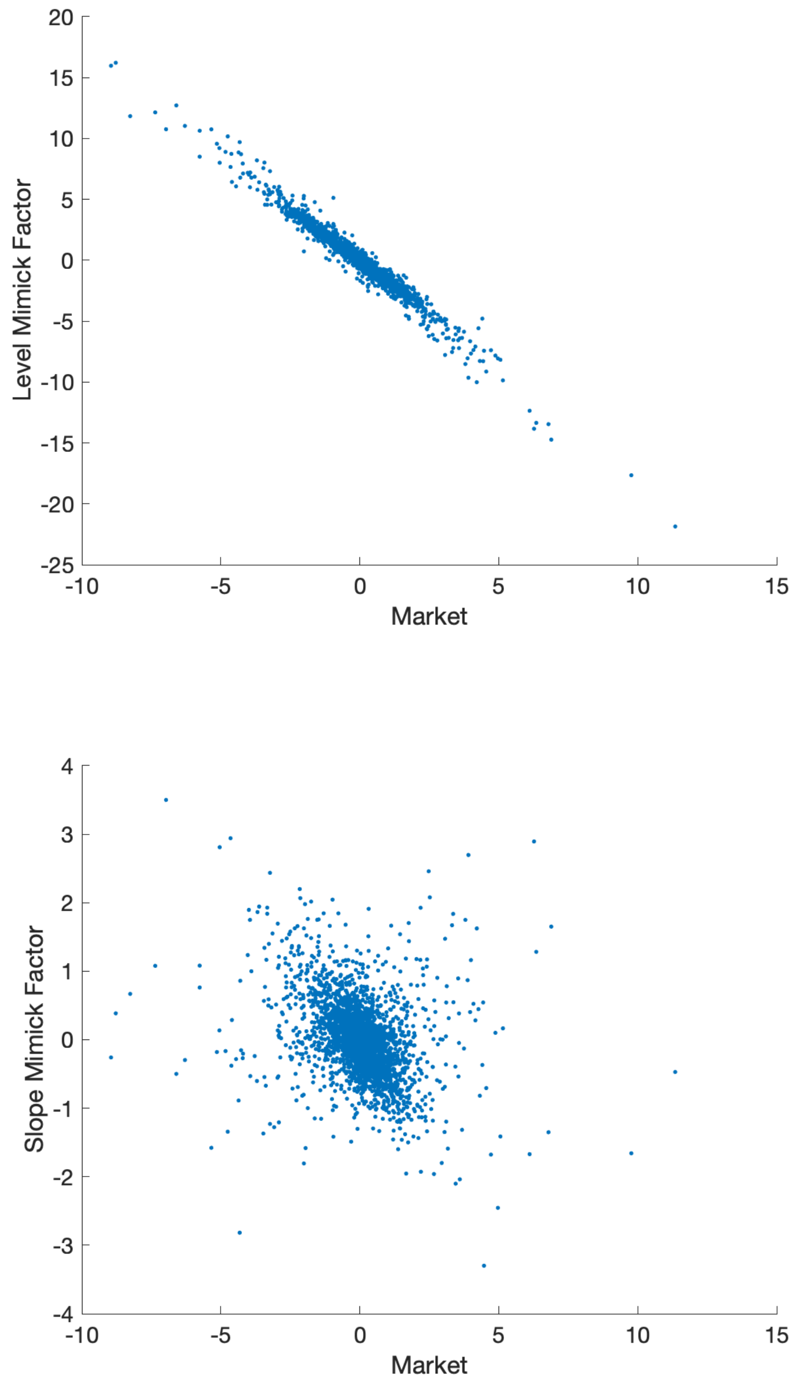
	Mrk	UMD	Port10	Crv
mean	9.15	0.33	11.92	6.09
std	19.83	15.83	23.82	12.57
Sharpe	0.46	0.02	0.50	0.48
corr Mrk	-	-	-	-
corr UMD	-35	-	-	-
corr Port10	86	5	-	-
corr Crv	86	0	93	-

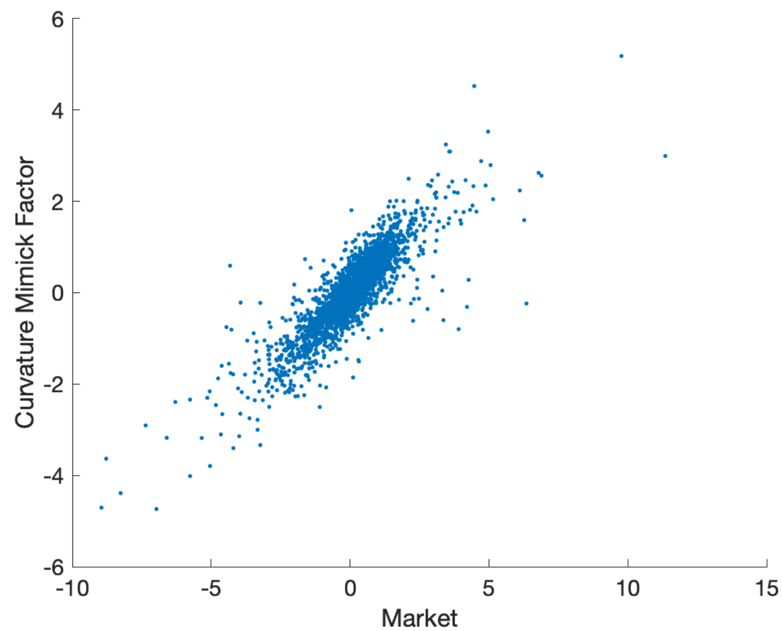
is higher than 90%. This is in line with the other results in the paper that show that curvature is positively loaded by high momentum portfolios. These correlation results show that the curvature mimicking portfolio and portfolio 10 are long market beta, while UMD is short market beta, in the sample. We also carried out our analysis at monthly frequencies. In that case, the curvature mimicking portfolio was less correlated with portfolio 10 (35%) and its Sharpe went up to 1.15. To put this into perspective, the monthly Sharpe ratios on market, UMD and portfolio 10 at monthly frequencies were 0.55, -0.02, and 0.54, respectively. However, due to the possibility of an extreme increase in error in variables in our analysis when using monthly frequencies - due to factor scores estimation - we preferred to use daily frequencies.

A.3.6.1 Scatter Plots of Factor mimicking Portfolios

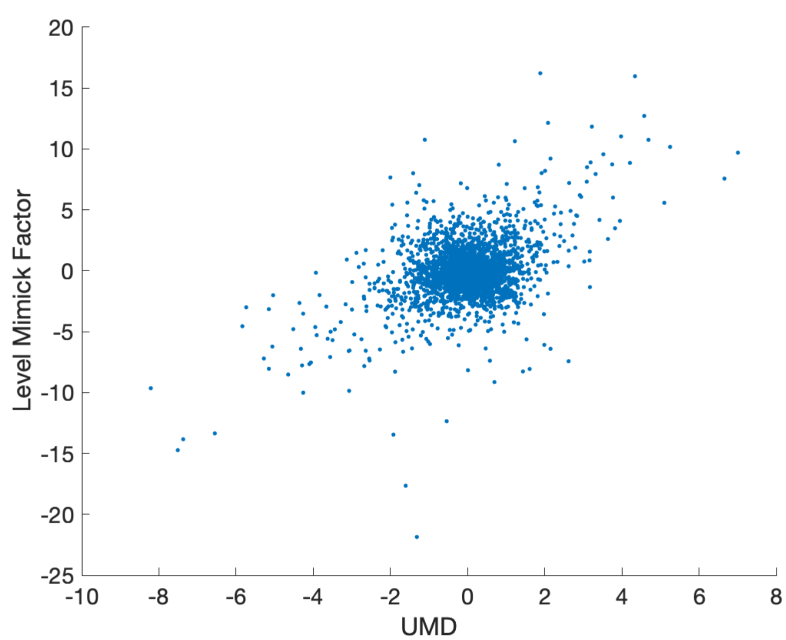
Figure A.3 reports various scatter plots to further highlight the relation of the factor mimicking portfolios with the market and UMD.

Figure A.3: Scatter Plots of Mimicking Portfolios and Factors.

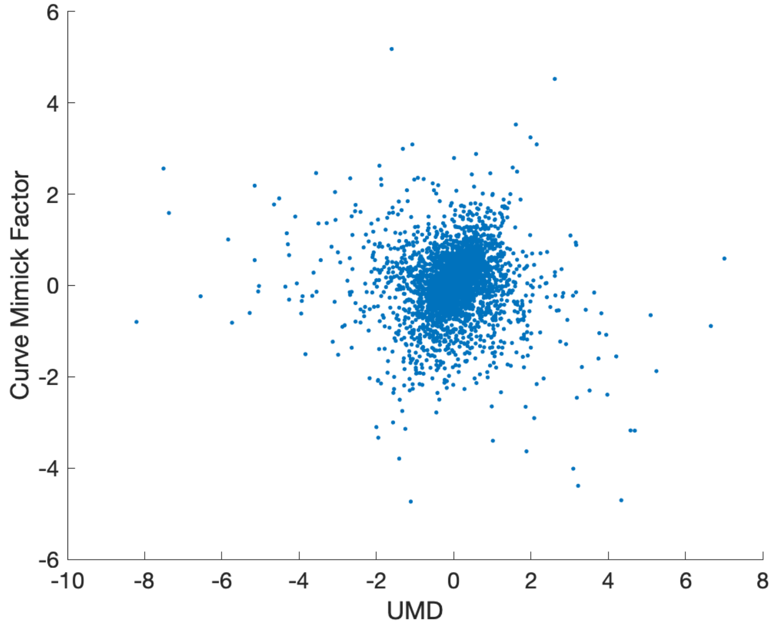


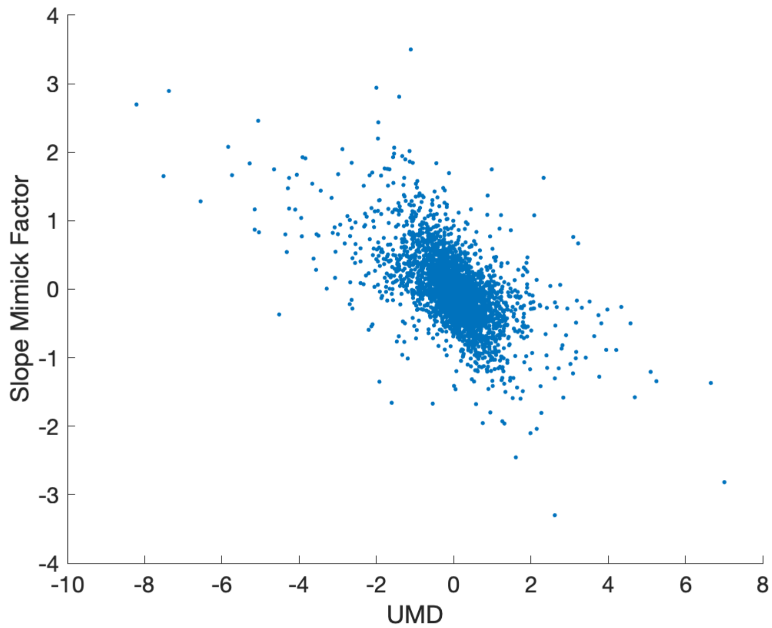


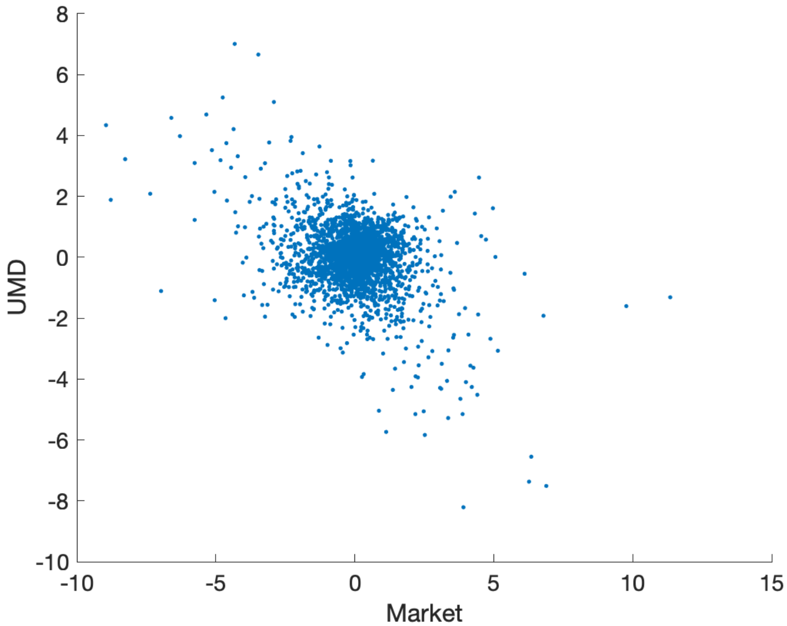
Curvature Mimick Factor
Market
6
4
2
0
-2
-4
-6
-10
-5
0
5
10
15



Level Mimick Factor
UMD
20
15
10
5
0
-5
-10
-15
-20
-25
-10
-8
-6
-4
-2
0
2
4
6
8


6
4
2
0
-2
-4
-6
Curve Mimick Factor
-10
-8
-6
-4
-2
0
2
4
6
8
UMD


4
3
2
1
0
-1
-2
-3
-4
Slope Mimick Factor
-10
-8
-6
-4
-2
0
2
4
6
8
UMD



A.4 Finite Sample Performance

To assess the performance of the theoretical results in a finite sample, we use simulations. We focus on a design relevant to the empirical results from Section 3. In particular, the starting point is the data from Section 3. We use the computed eigenfunctions and factor scores as if they were the population ones. We also use the estimated loadings from the times series regression to compute residuals. The time series regression is the one of returns on the portfolios regressed on market excess return and the three factor scores. We then simulate error terms and factors ensuring that the same structure of the original data is preserved. The details on how the data are simulated is given in Algorithm 3. Our results use 1000 simulations to compute the quantities of interest with increasing sample size $n_0 \in \{1250, 2500, 5000\}$ and varying levels of noise relative to the original data to assess the sensitivity of the results.

Algorithm 3 Monte Carlo Simulation.

Define

$X_t = \left(R^e_{t,m}, \hat{S}_{t,1}, \hat{S}_{t,2}, \hat{S}_{t,3}\right)'$ where $\hat{S}_{t,1}, \hat{S}_{t,2}, \hat{S}_{t,3}$ are the first three factor scores estimated from the sample data in the empirical section of the paper;

$\left\{\hat{\Phi}_i(\cdot) : i = 1, 2, 3\right\}$ to be the estimated eigenfunctions;

Compute

The mean and covariance matrix of $\{X_t\}$, say $\hat{\mu}_X$, $\hat{C}_X$;

The estimator $\left(\hat{a}_0, \hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3\right)$ (of dimension 11×4) from the time series regression $R_t = a_0 + \beta_0 R^e_{t,m} + \sum_{i=1}^{3} \beta_i \hat{S}_{t,i} + \varepsilon_t$;

The residuals from this regression and estimate their covariance matrix, say $\hat{C}_\varepsilon$.

Set

$n_0 \in \{1250, 2500, 5000\}$, $\tau \in \{1/2, 1, 2\}$.

For $r = 1, 2, ..., 1000$

Simulate error terms $\left\{\varepsilon^{[r]}_t : t = 1, 2, ..., n_0\right\}$ from a multivariate normal distribution with mean zero and covariance $\tau\hat{C}_\varepsilon$;

Simulate factor scores $\left\{X^{[r]}_t : t = 1, 2, ..., n_0\right\}$ from a multivariate normal distribution with mean $\hat{\mu}_X$ and covariance $\hat{C}_X$;

Define $\left\{R^{[r]}_t : t = 1, 2, ..., n_0\right\}$ where $R^{[r]}_t = \hat{a}_0 + \sum_{i=1}^{4} \hat{\beta}_{i-1} X^{[r]}_{t,i} + \varepsilon^{[r]}_t$;

Define $\left\{F^{[r]}_t : t = 1, 2, ..., n_0\right\}$ where $F^{[r]}_t(\cdot) = \sum_{i=1}^{3} X^{[r]}_{t,i+1} \hat{\Phi}_i(\cdot)$;

Carry out the estimations as in the paper but using the simulated data;

End of for loop.

In Algorithm 3 we use the superscript r within square brackets to mean that the variables are the result of the r^{th} simulation. This should not be confused with the superscripts used to denote the k risky asset of the l factor. Hence, in Algorithm 3, $R^{[r]}_t$ is the 11×1 vector of risky assets from the r^{th} simulation.

In order to understand the results, note that we use τ in the simulation Algorithm 3 to vary the signal to noise. In particular, $\tau = .5$ corresponds to a signal to noise ratio increased by two times relative to the sample data in the empirical section. On the other hand, $\tau = 2$ implies a reduction of the signal to noise by two. The results show that a relatively large sample size may be needed when the signal to noise is very low. Consistent estimation of the risk premia for very low signal to noise (i.e. $\tau = 2$) can be challenging. The detailed results are in Tables A.10, A.11, A.12, and A.13. To summarise these results, we note that the estimation of the loadings is not very sensitive to the level of noise. However, the risk premium estimation is more challenging and does require larger sample sizes when the signal to noise drops substantially.

Table A.10: Time Series Regression: Estimated loadings. The true values are reported on the top line. Mean and s.e. are the mean and standard errors from from 1000 simulations.

			Const.	Market	Level	Slope	Curvature
n_0	τ	True	0.008	1.092	0.428	-2.995	1.994
1250	0.5	mean	0.008	1.092	0.424	-2.986	2.000
1250	0.5	s.e.	0.000	0.001	0.003	0.010	0.020
2500	0.5	mean	0.008	1.092	0.427	-2.988	2.011
2500	0.5	s.e.	0.000	0.000	0.002	0.007	0.014
5000	0.5	mean	0.008	1.093	0.427	-2.981	1.991
5000	0.5	s.e.	0.000	0.000	0.001	0.005	0.010
1250	1	mean	0.008	1.092	0.423	-2.983	2.002
1250	1	s.e.	0.001	0.001	0.004	0.014	0.028
2500	1	mean	0.008	1.092	0.427	-2.985	2.018
2500	1	s.e.	0.000	0.001	0.003	0.010	0.019
5000	1	mean	0.008	1.093	0.427	-2.975	1.990
5000	1	s.e.	0.000	0.000	0.002	0.007	0.014
1250	2	mean	0.008	1.092	0.421	-2.979	2.005
1250	2	s.e.	0.001	0.001	0.006	0.020	0.040
2500	2	mean	0.008	1.092	0.426	-2.981	2.028
2500	2	s.e.	0.001	0.001	0.004	0.014	0.027
5000	2	mean	0.008	1.093	0.427	-2.967	1.989
5000	2	s.e.	0.000	0.001	0.003	0.010	0.020

Table A.11: Cross-sectional Regression: Estimated risk premia from the two-step regression. The true values are reported on the top line. Mean and s.e. are the mean and standard errors from from 1000 simulations. Numbers multiplied by 100.

			Const.	Market	Level	Slope	Curvature
n_0	τ	True	-0.016	3.918	-2.584	0.124	0.800
1250	0.5	mean	-0.025	3.931	-2.777	0.020	0.655
1250	0.5	s.e.	0.015	0.116	0.066	0.017	0.021
2500	0.5	mean	-0.018	3.975	-2.696	0.064	0.708
2500	0.5	s.e.	0.010	0.081	0.045	0.012	0.017
5000	0.5	mean	-0.021	3.953	-2.643	0.095	0.760
5000	0.5	s.e.	0.007	0.058	0.035	0.008	0.011
1250	1	mean	-0.022	3.929	-2.880	-0.052	0.538
1250	1	s.e.	0.022	0.117	0.091	0.024	0.028
2500	1	mean	-0.019	3.972	-2.791	0.013	0.625
2500	1	s.e.	0.014	0.082	0.062	0.016	0.022
5000	1	mean	-0.023	3.953	-2.694	0.067	0.719
5000	1	s.e.	0.010	0.059	0.048	0.012	0.015
1250	2	mean	0.014	3.894	-2.930	-0.134	0.388
1250	2	s.e.	0.036	0.121	0.131	0.034	0.034
2500	2	mean	-0.010	3.955	-2.907	-0.057	0.501
2500	2	s.e.	0.021	0.083	0.088	0.022	0.028
5000	2	mean	-0.025	3.950	-2.777	0.017	0.642
5000	2	s.e.	0.014	0.060	0.067	0.017	0.021

Table A.12: Risk Price from GMM: Estimated risk prices. The true values are reported on the top line. Mean and s.e. are the mean and standard errors from from 1000 simulations.

			Market	Level	Slope	Curvature
n_0	τ	True	-0.280	-1.303	0.087	16.925
1250	0.5	mean	-0.252	-1.220	-0.487	13.247
1250	0.5	s.e.	0.006	0.027	0.084	0.350
2500	0.5	mean	-0.261	-1.248	-0.258	14.696
2500	0.5	s.e.	0.004	0.018	0.058	0.281
5000	0.5	mean	-0.270	-1.274	-0.090	15.828
5000	0.5	s.e.	0.003	0.014	0.042	0.191
1250	1	mean	-0.229	-1.148	-0.820	10.626
1250	1	s.e.	0.008	0.036	0.113	0.447
2500	1	mean	-0.249	-1.213	-0.510	12.847
2500	1	s.e.	0.006	0.025	0.079	0.370
5000	1	mean	-0.265	-1.262	-0.241	14.832
5000	1	s.e.	0.004	0.019	0.058	0.259
1250	2	mean	-0.191	-1.006	-1.090	7.437
1250	2	s.e.	0.010	0.049	0.149	0.533
2500	2	mean	-0.226	-1.142	-0.826	10.195
2500	2	s.e.	0.008	0.035	0.107	0.464
5000	2	mean	-0.252	-1.228	-0.500	12.990
5000	2	s.e.	0.006	0.026	0.080	0.342

Table A.13: Risk Premia from GMM: Estimated risk premia based on GMM. The true values are reported on the top line. Mean and s.e. are the mean and standard errors from from 1000 simulations. Numbers multiplied by 100.

			Market	Level	Slope	Curvature
n_0	τ	True	3.861	-3.059	0.189	0.844
1250	0.5	mean	3.852	-2.869	0.079	0.577
1250	0.5	s.e.	0.114	0.061	0.016	0.019
2500	0.5	mean	3.907	-2.993	0.120	0.681
2500	0.5	s.e.	0.080	0.042	0.012	0.016
5000	0.5	mean	3.883	-3.035	0.152	0.763
5000	0.5	s.e.	0.057	0.032	0.008	0.011
1250	1	mean	3.868	-2.719	0.030	0.431
1250	1	s.e.	0.114	0.080	0.021	0.024
2500	1	mean	3.915	-2.941	0.074	0.563
2500	1	s.e.	0.080	0.056	0.016	0.020
5000	1	mean	3.890	-3.013	0.121	0.691
5000	1	s.e.	0.058	0.043	0.011	0.015
1250	2	mean	3.886	-2.434	-0.003	0.280
1250	2	s.e.	0.114	0.105	0.027	0.028
2500	2	mean	3.921	-2.809	0.024	0.409
2500	2	s.e.	0.080	0.076	0.020	0.024
5000	2	mean	3.895	-2.956	0.072	0.575
5000	2	s.e.	0.058	0.058	0.015	0.019

References

[1] Aldous, D. (1976) A Characterization of Hilbert Space Using the Central Limit Theorem. Journal of the London Mathematical Society 14, 376-380.

[2] Bathia, R. (1997) Matrix Analysis. New York: Springer.

[3] Breeden, D.T (1979) An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities. Journal of Financial Economics 7, 265-296

[4] Breeden, DT., Gibbons, M.R., and Litzenberger, R. (1989) Empirical Tests of the Consumption-Oriented CAPM. Journal of Finance 44, 231-262.

[5] Bosq, D. (2000). Linear Processes in Function Spaces: Theory and Applications, volume 149 of Lecture Notes in Statistics.

[6] Carhart, M.M. (1997) On Persistence in Mutual Fund Performance. The Journal of Finance 52, 57-82.

[7] Dauxois, J., Pousse, A., and Romain, Y., (1982) Asymptotic Theory for the Principal Component Analysis of a Vector Random Function: Some Applications to Statistical Inference. Journal of Multivariate Analysis 12, 136-154.

[8] De Vito, E., Umanitá V., and Villa S. (2013) An Extension of Mercer Theorem to Matrix-Valued Measurable Kernels. Applied and Computational Harmonic Analysis 34, 339-351.

[9] Horváth, L., and Kokoszka P. (2012) Inference for Functional Data With Applications. Berlin: Springer.

[10] Huberman, G., Kandel, S., and Stambaugh, R.F. (1987) Mimicking Portfolios and Exact Arbitrage Pricing. Journal of Finance 42, 1-9.