Preliminary Draft
Please Do Not Circulate
Comments Welcome

Measuring Horizon-Specific Systematic Risk via Spectral Betas

Federico M. Bandi*, Shomesh E. Chaudhuri†, Andrew W. Lo‡, Andrea Tamoni§

June 15, 2018

Abstract

Frequency is a key dimension of risk. Using a suitable multivariate Wold representation, we introduce a notion of frequency-specific or *spectral beta*. We show that the traditional beta can be expressed as a linear combination of spectral betas without cross-beta terms. Hence, all frequency-specific information is contained in the spectral betas. We discuss how spectral betas may be estimated using either nonparametric or parametric methods relying on the extraction of the Wold coefficients. In both cases, the representations and the methods are akin to linear projections on moving averages *in the time domain*. Thus, interpretation and applicability are facilitated with respect to frequency-domain analogues. Using NYSE and AMEX stock returns from 1972 to 2016, we show that measuring frequency-dependent systematic risk through spectral betas leads to portfolios with lower out-of-sample variance relative to specifications in which risk is, as generally assumed, constant across frequencies. We also show that a CAPM model in which systematic risk is captured by a low-frequency spectral (market) beta outperforms well-known multi-factor specifications, thereby achieving dimension reduction as well as economic interpretability.

Keywords: systematic risk, factor models, portfolio optimization, cross-sectional pricing.

JEL Classification: C22, C32, E32; G11; G12

---

*Carey Business School, Johns Hopkins University and Edhec-Risk Institute.
†Electrical Engineering and Computer Science Department, MIT.
‡Laboratory for Financial Engineering, Sloan School of Management, MIT; Electrical Engineering and Computer Science Department, MIT; Computer Science and Artificial Intelligence Laboratory, MIT; Santa Fe Institute.
§Finance Department, London School of Economics.

Contents

1 Introduction

Systematic risk is typically modeled as being constant across frequencies. While this assumption has rarely been questioned, it is indeed an "assumption," one which imposes potentially strong restrictions on risk measurements across horizons.

Beginning with the work of Hannan (1963a; 1963b), methods of inference for regression slopes in the frequency domain have provided a framework to identify frequency-specific betas. In the frequency domain, spectral and co-spectral power, calculated using Fourier transforms, offer a natural way to study the frequency-specific components of variance and covariance. Specifically, spectral power permits identification of the variability of a time series resulting from fluctuations at a specific frequency while co-spectral power measures co-variation over specific frequencies. When the signals are in phase at a given frequency (i.e., the "peaks" and the "valleys" of the assumed time series coincide), the co-spectral power is positive at that frequency. When they are out of phase, it is negative. The ratio between co-spectral and spectral power yields a natural frequency-specific beta estimate, one which is defined in *the frequency domain*. Reliance on the frequency domain, however, renders interpretation and applicability not entirely straightforward. This observation justifies the, arguably, very limited use of frequency-domain methods in economics and finance.[1]

The goal of this paper is to provide frequency-specific assessments of systematic risk in *the time domain*. Our approach begins with a classical Wold (1938) representation, which we then modify to derive an equivalent (orthogonal) representation in which a multivariate covariance-stationary time series is modeled as an infinite sum of orthogonal frequency-specific Wold representations. Each frequency-specific Wold representation defines a (Wold) component of the original series operating over a certain horizon.

We define the *spectral beta* as the beta between components of two time series with cycles of equal periodicity, i.e., the beta between corresponding Wold components. Exploiting the orthogonal Wold representation, we also represent the overall beta between two time series as a weighted average of spectral betas in which the weights depend on the informational content (i.e., the variance) of the corresponding Wold components. Because of the orthogonality of the Wold components, no cross-beta terms appear in the representation of the overall beta, thereby guaranteeing that all frequency-specific information is captured by the corresponding spectral beta.

We study the identification of the Wold components and spectral betas using parametric methods (which explicitly rely on the Wold representation of the original covariance-stationary multivariate process) and nonparametric methods. We further discuss the logical

[1]Berkowitz (2001), Cogley (2001) and Dew-Becker and Giglio (2016) constitute notable exceptions.

connections between the two approaches, their relative benefits, and their empirical performance.

Economically, our goal is to add a dimension to the assessment of systematic risk. In essence, the magnitude of systematic risk is not just a function of investment style, it is also a function of frequency. After dividing stocks into size and value quintiles, we show that the market spectral betas increase (almost) monotonically for small and value stocks when going from high to low frequencies. Conversely, they decrease almost monotonically for large and growth stocks, again when going from high to low frequencies. The reported spread in the low frequency betas—with high betas associated with value stocks and low betas associated with growth stocks—are suggestive of a frequency-based justification of the ubiquitous value premium in which low-frequency dynamics affect the valuation of cross-sectional returns at higher frequencies.

In light of these findings, we explore pricing across frequencies using spectral betas. For the 25 Fama-French portfolios, we find that the pricing errors are substantially reduced as we move from quantifying risk using high-frequency betas to quantifying risk using low-frequency betas. Given the structure of our assumed model (which employs only one frequency at a time), a decrease in alpha, as the time horizon increases, is suggestive of the role of frequency in the assessment of systematic risk. Indeed, a CAPM model based on low-frequency spectral betas performs better than the three Fama-French factor model. This result, in turn, suggests that the horizon of the risk exposures may provide an important dimension to reduce the dimensionality of pricing models, the latter being the focus of much existing work.

In addition, the identification of the particular horizon responsible for a given strategy's expected return and volatility can give investors a supplementary lever with which to manage the risk/reward profile of a portfolio. For a specific set of historical asset returns, we formulate a *spectral* factor model which decomposes the return covariance matrix into the sum of high, medium, and low-frequency components. The decomposition facilitates the investor's decision regarding best and worst sources of diversification at different frequencies, thereby leading to superior portfolio allocations. We assess the economic significance of the proposed spectral betas by evaluating the out-of-sample performance of alternative (optimally-designed) covariance matrices of stock returns. For NYSE and AMEX stocks sampled between 1972 to 2016, using spectral analogues to traditional factor models leads to systematically lower out-of-sample variability of the resulting portfolios.

In order to motivate the association between risk and frequency, and further illustrate it in this Introduction by way of a more anecdotal example, we plot in Fig. 1 the Berkshire Hathaway's spectral betas with respect to both the value- and the equally-weighted market (monthly) returns. Consistent with intuition, and several corporate statements over the years, Warren Buffett's strategy amounts to a long-term value investing approach. The betas increase as a function of the time horizon.

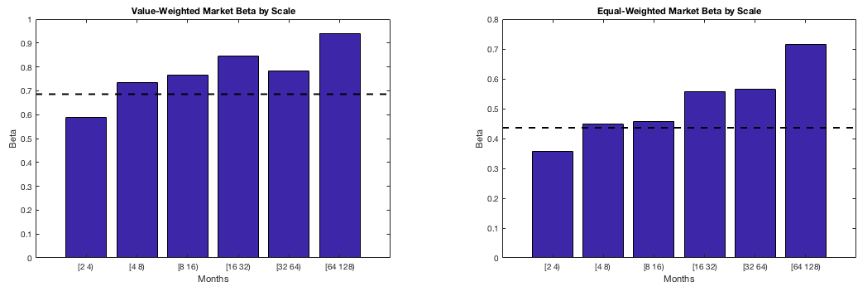

Figure 1: Decomposition of Buffett's market beta into its various frequency components. The $x-$axis displays the frequency cycles, e.g. the first bar captures the beta corresponding to cycles with length between 2^1 and 2^2 months, the second corresponds to cycles with length between 2^2 and 2^4 months, and so on.

The paper proceeds as follows. In Section 2.1, we review how to transition from classical Wold representations to *extended* Wold representations in which a multivariate time series is expressed as an infinite sum of Wold components operating over different frequencies. Section 2.2 introduces the notion of spectral beta and formalizes the representation of the overall beta in terms of a suitably-defined weighted average of spectral betas. In Sections 2.3 and 2.4, we study parametric and nonparametric extraction of the Wold components and the resulting spectral betas. Section 3 verifies the empirical validity of both estimation strategies. Portfolio optimization using spectral factor models relying on spectral betas is contained in Section 4. In Section 5, we return to the economics of spectral betas and their potential for dimensionality reductions in cross-sectional pricing. Section 6 concludes.

2 Representation and Identification

We begin with the description of a helpful (for the purpose of zooming into components with different cycles) Wold representation of a multivariate covariance-stationary time series. Without loss of generality, we focus on the bivariate case henceforth to present ideas without complicating notation unnecessarily. After reviewing the multivariate Wold representation, we apply it to construct a novel representation of the beta of a given bivariate time series as a linear combination of the betas of its Wold (frequency-specific) components. We then discuss parametric and nonparametric identification issues.

2.1 Multivariate Wold Representation

Let $\mathbf{x} = \{(x_t^1, x_t^2)^\intercal\}_{t\in\mathbb{Z}}$ be a covariance-stationary bivariate process defined onto the space $L^2(\Omega, \mathcal{F}, \mathbb{P})$ with the usual properties. For simplicity, assume the bivariate process is mean zero. Adding a mean would, of course, simply add a constant term to its Wold representation (below).

Define the white noise process $\boldsymbol{\varepsilon} = \{(\varepsilon_t^1, \varepsilon_t^2)^\intercal\}_{t\in\mathbb{Z}}$ such that $\mathbb{E}[\boldsymbol{\varepsilon}] = 0$ and $\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\intercal] = \Sigma_\varepsilon$, where Σ_ε is a covariance matrix of dimension 2×2. For any t in $\mathbb{Z}$, $\mathbf{x}$ satisfies the following Wold representation:

$$\begin{pmatrix} x_t^1 \\ x_t^2 \end{pmatrix} = \sum_{k=0}^{\infty} \begin{pmatrix} \alpha_k^1 & \alpha_k^2 \\ \alpha_k^3 & \alpha_k^4 \end{pmatrix} \begin{pmatrix} \varepsilon_{t-k}^1 \\ \varepsilon_{t-k}^2 \end{pmatrix} = \sum_{k=0}^{\infty} \boldsymbol{\alpha}_k \boldsymbol{\varepsilon}_{t-k}, \tag{1}$$

with $\sum_{k=0}^{\infty} \mathrm{tr}^{1/2}(\boldsymbol{\alpha}_k^\intercal \boldsymbol{\alpha}_k) < \infty$ and $\boldsymbol{\alpha}_0 = I_2$, where the equality is in the L^2-norm.

Straightforward aggregation of the system's shocks now leads to the equivalent *extended* Wold representation:[2]

$$\begin{pmatrix} x_t^1 \\ x_t^2 \end{pmatrix} = \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k^{(j)} \boldsymbol{\varepsilon}_{t-k2^j}^{(j)} = \sum_{j=1}^{\infty} \mathbf{x}_t^{(j)} \tag{2}$$

in which, for any $j \in \mathbb{N}$, the 2×2 matrices $\boldsymbol{\Psi}_k^{(j)}$ are the unique discrete Haar transforms (DHT) of the original Wold coefficients, i.e.,

$$\boldsymbol{\Psi}_k^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \boldsymbol{\alpha}_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \boldsymbol{\alpha}_{k2^j+2^{j-1}+i} \right), \tag{3}$$

and the 2×1 vectors $\boldsymbol{\varepsilon}_t^{(j)}$ are the DHTs of the original Wold shocks, i.e.,

$$\boldsymbol{\varepsilon}_t^{(j)} = \frac{1}{\sqrt{2^j}} \left(\sum_{i=0}^{2^{j-1}-1} \boldsymbol{\varepsilon}_{t-i} - \sum_{i=0}^{2^{j-1}-1} \boldsymbol{\varepsilon}_{t-2^{j-1}-i} \right). \tag{4}$$

[2] We refer the interested reader to Bandi, Perron, Tamoni, and Tebaldi (2017) and to Ortu, Severino, Tamoni, and Tebaldi (2017) for a formal justification based on Hilbert space theory.

The representation in Eq. (2) amounts to an orthogonal decomposition of the original bivariate time series into uncorrelated components (for every t) operating over alternative frequencies (denoted by the index j). This is easy to see. Take the first and second frequencies, i.e., $j = 1$ and $j = 2$, and notice that the corresponding shocks are defined as follows:

$$\boldsymbol{\varepsilon}_t^{(1)} = \begin{pmatrix} \frac{\varepsilon_t^1 - \varepsilon_{t-1}^1}{\sqrt{2}} \\ \frac{\varepsilon_t^2 - \varepsilon_{t-1}^2}{\sqrt{2}} \end{pmatrix}, \quad \boldsymbol{\varepsilon}_{t-2}^{(1)} = \begin{pmatrix} \frac{\varepsilon_{t-2}^1 - \varepsilon_{t-3}^1}{\sqrt{2}} \\ \frac{\varepsilon_{t-2}^2 - \varepsilon_{t-3}^2}{\sqrt{2}} \end{pmatrix}, \quad \ldots$$

and

$$\boldsymbol{\varepsilon}_t^{(2)} = \begin{pmatrix} \frac{(\varepsilon_t^1 + \varepsilon_{t-1}^1) - (\varepsilon_{t-2}^1 + \varepsilon_{t-3}^1)}{\sqrt{4}} \\ \frac{(\varepsilon_t^2 + \varepsilon_{t-1}^2) - (\varepsilon_{t-2}^2 + \varepsilon_{t-3}^2)}{\sqrt{4}} \end{pmatrix}, \quad \boldsymbol{\varepsilon}_{t-4}^{(2)} = \begin{pmatrix} \frac{(\varepsilon_{t-4}^1 + \varepsilon_{t-5}^1) - (\varepsilon_{t-6}^1 + \varepsilon_{t-7}^1)}{\sqrt{4}} \\ \frac{(\varepsilon_{t-4}^2 + \varepsilon_{t-5}^2) - (\varepsilon_{t-6}^2 + \varepsilon_{t-7}^2)}{\sqrt{4}} \end{pmatrix}, \quad \ldots$$

The coefficient matrices have the same structure, e.g., for $j = 1$ we have

$$\boldsymbol{\Psi}_0^{(1)} = \begin{pmatrix} \frac{\alpha_0^1 - \alpha_1^1}{\sqrt{2}} & \frac{\alpha_0^2 - \alpha_1^2}{\sqrt{2}} \\ \frac{\alpha_0^3 - \alpha_1^3}{\sqrt{2}} & \frac{\alpha_0^4 - \alpha_1^4}{\sqrt{2}} \end{pmatrix}, \quad \boldsymbol{\Psi}_1^{(1)} = \begin{pmatrix} \frac{\alpha_2^1 - \alpha_3^1}{\sqrt{2}} & \frac{\alpha_2^2 - \alpha_3^2}{\sqrt{2}} \\ \frac{\alpha_2^3 - \alpha_3^3}{\sqrt{2}} & \frac{\alpha_2^4 - \alpha_3^4}{\sqrt{2}} \end{pmatrix}, \quad \ldots$$

and, for $j = 2$,

$$\boldsymbol{\Psi}_0^{(2)} = \begin{pmatrix} \frac{(\alpha_0^1 + \alpha_1^1) - (\alpha_2^1 + \alpha_3^1)}{\sqrt{4}} & \frac{(\alpha_0^2 + \alpha_1^2) - (\alpha_2^2 + \alpha_3^2)}{\sqrt{4}} \\ \frac{(\alpha_0^3 + \alpha_1^3) - (\alpha_2^3 + \alpha_3^3)}{\sqrt{4}} & \frac{(\alpha_0^4 + \alpha_1^4) - (\alpha_2^4 + \alpha_3^4)}{\sqrt{4}} \end{pmatrix}, \quad \boldsymbol{\Psi}_1^{(2)} = \begin{pmatrix} \frac{(\alpha_4^1 + \alpha_5^1) - (\alpha_6^1 + \alpha_7^1)}{\sqrt{4}} & \frac{(\alpha_4^2 + \alpha_5^2) - (\alpha_6^2 + \alpha_7^2)}{\sqrt{4}} \\ \frac{(\alpha_4^3 + \alpha_5^3) - (\alpha_6^3 + \alpha_7^3)}{\sqrt{4}} & \frac{(\alpha_4^4 + \alpha_5^4) - (\alpha_6^4 + \alpha_7^4)}{\sqrt{4}} \end{pmatrix}, \quad \ldots$$

It is now straightforward to verify that the shocks associated with the first and second components are white noise processes (with variances Σ_ε) when defined on the supports $S_t^{(1)} = \{t - k2 : k \in \mathbb{Z}\}$ and $S_t^{(2)} = \{t - k2^2 : k \in \mathbb{Z}\}$. In addition, the two components are orthogonal at all leads and lags. More generally, for generic components j and k, we have

$$\mathbb{E}\left[\mathbf{x}_{t-m2^j}^{(j)} {\mathbf{x}_{t-n2^l}^{(k)}}^\intercal\right] = \mathbf{0} \qquad \forall j \neq k, \quad \forall m, n \in \mathbb{N}_0, \quad \forall t \in \mathbb{Z}. \tag{5}$$

In essence, the bivariate time series $\mathbf{x}$ has been re-expressed as an infinite sum of Wold components, i.e., $\mathbf{x}_t = \sum_{j=1}^{\infty} \mathbf{x}_t^{(j)}$. Each Wold component has a Wold representation with white noise shocks suitably defined on an enlarging support $S_t^{(j)} = \{t - k2^j : k \in \mathbb{Z}\}$.

Therefore, the generic j-th Wold components will capture cycles with length between 2^{j-1} and 2^j periods.

2.2 *Spectral* β: A Novel Representation

Linear factor models are routinely used in financial applications, like market model regressions, the capital asset pricing model (CAPM) introduced by Sharpe (1964) and Lintner (1965), arbitrage pricing theory, and the Fama-French (1992) three-factor model, to name just a few. However, the estimated beta coefficients in these models are "static" measures unable to capture horizon-specific dependencies between variables.

Spectral regression, as proposed by Hannan (1963a) and Hannan (1963b), on the other hand, does capture the sensitivity of the dependent variable to fluctuations in the independent variable(s) over different frequencies. Its reliance on frequency domain, however, does not facilitate interpretation and/or applicability.

In what follows, we employ the extended Wold representation in the previous section, along with its properties, to introduce a novel representation of the overall beta of the time series x_t^1 and x_t^2 in $\mathbf{x}$ as a linear combination of betas on the Wold components. The beta associated with corresponding (i.e., same frequency j) Wold components measures horizon-specific dependence between variables and defines our notion of *spectral beta*.

Theorem 1 (A β representation.) *Assume* $\mathbf{x} = \{(x_t^1, x_t^2)^\intercal\}_{t\in\mathbb{Z}}$ *satisfies Eq.* (1). *Define the spectral beta associated with frequency* j *as* $\beta^{(j)} = \frac{\mathbb{E}\left[x_t^{1,(j)},\ x_t^{2,(j)}\right]}{\mathbb{V}\left[x_t^{2,(j)}\right]}$. *The overall beta would, therefore, conform with*

$$\beta = \frac{\mathbb{C}\left[x_t^1, x_t^2\right]}{\mathbb{V}\left[x_t^2\right]} = \sum_{j=1}^{\infty} v^{(j)}\beta^{(j)}, \tag{6}$$

where $v^{(j)} = \frac{\mathbb{V}\left[x_t^{2,(j)}\right]}{\mathbb{V}\left[x_t^2\right]}$.

Proof. *Immediate, given the representation in Eq.* (2) *and its properties.*

In other words, in light of the extended Wold representation, the classical beta can be expressed as a weighted average of spectral betas with weights directly related to the *relative* informational content of the corresponding frequency. The latter is, of course, defined by $v^{(j)} = \frac{\mathbb{V}\left[x_t^{2,(j)}\right]}{\mathbb{V}\left[x_t^2\right]}$.

We note that one could write the spectral betas, and the overall beta, directly in terms of the Wold parameters. In this paper, we filter the components, something which is of independent interest, before computing spectral betas using empirical covariances and variances. Since we do not exploit the structure of the betas in terms of Wold parameters directly, we refer the interested reader to the paper's Appendix for the corresponding result.

We now turn to the identification of the components. We do so using both parametric and nonparametric methods. The empirical validity of the resulting spectral covariances and spectral betas is assessed in Section 3.

2.3 Parametric Identification

In order to operationalize the extended Wold representation in Eq. (2), we first need to compute the classical Wold coefficients, $\boldsymbol{\alpha}_k$.

To this end, we assume that the bivariate time series of interest, $\mathbf{x}_t = (x_t^1, x_t^2)^\intercal$, follows a linear vector autoregressive (VAR) process of order p (VAR(p)) of the form:

$$\mathbf{x}_t = A_1 \mathbf{x}_{t-1} + \ldots + A_p \mathbf{x}_{t-p} + \varepsilon_t, \tag{7}$$

where the A_is, with $i = 1, \ldots, p$, are 2×2 parameter matrices and the error process, $\varepsilon_t = (\varepsilon_t^1, \varepsilon_t^2)^\intercal$, is a 2-dimensional zero-mean white noise process with covariance matrix $\mathbb{E}(\varepsilon_t \varepsilon_t^\intercal) = \Sigma_\varepsilon$.

Without loss of generality, we focus on the VAR(1) case since any 2-dimensional VAR(p) process, as in Eq. (7), can be written as a $2p$-dimensional VAR(1) process by stacking p consecutive $\mathbf{x}_t$ variables in a $(2p \times 1)$-dimensional vector $\mathbf{X}_t = \left(\mathbf{x}_t^\intercal, \ldots, \mathbf{x}_{t-p+1}^\intercal\right)^\intercal$. Thus, we have

$$\mathbf{X}_t = A\mathbf{X}_{t-1} + U_t, \tag{8}$$

where the matrix A is sometimes referred to as the *companion* matrix of the VAR(p) process. It is well-known that $\mathbf{X}_t$ is stable if

$$\det\left(I_{2p} - Az\right) \neq 0 \quad \forall z \in \mathcal{C}, \ |z| \leq 1.$$

It is also easy to see that this condition is equivalent to assuming that all of the eigenvalues of A have modulus less than one.

Under standard assumptions, a stable VAR has time invariant means and a time-invariant variance/covariance matrix, and is, therefore, covariance stationary (see, e.g., Lutkepohl (2005), Chapter 2). Said differently, the process admits a Wold representation. Indeed, the stability of the process ensures the existence of the inverse VAR operator, $A(L)^{-1} = \sum_{k=0}^{\infty} \boldsymbol{\alpha}_k L^k$, where $\boldsymbol{\alpha}_k = A^k$. As a result, one can obtain the following Wold moving average representation of $\mathbf{X}_t$:

$$\mathbf{X}_t = \sum_{k=0}^{\infty} \boldsymbol{\alpha}_k U_{t-k}. \tag{9}$$

Given this discussion, parametric identification of the Wold components proceeds as follows. We assume a VAR(p) process for $\mathbf{x}_t$, obtain its companion representation in Eq. (8), and estimate the resulting VAR(1) model using unrestricted least squares. We then obtain the Wold coefficients using the restriction, $\boldsymbol{\alpha}_k = A^k$, and plug them into Eq. (3) to evaluate the coefficients of the extended Wold representation. Finally, using the residuals from the VAR(1) along with Eq. (4), we obtain the frequency-specific innovations of the extended Wold representation.

Parametric identification implies reliance on a linear (covariance-stationary) model. Even though we will show that linearity is not inconsistent with our data, we now turn to an identification procedure which, by construction, dispenses with linearity.

2.4 Nonparametric Identification

In this subsection, we discuss how to use low-pass filters in the time domain to calculate the frequency components of a covariance-stationary bivariate process. Similar to Beveridge and Nelson (1981), who popularized the decomposition of a nonstationary time series into a trend component and a transient component, we decompose a covariance-stationary process into components (more than 2, in our case) operating over different frequencies and with heterogeneous levels of persistence. The filtered components will subsequently be used to identify the component-wise or spectral betas. Because the procedure does not require estimation of the Wold parameters and hinges solely on suitably-selected moving averages of the available data, we characterize it as being *nonparametric* in nature.

Given a time series $\{x_t\}_{t\in\mathbb{Z}}$, we construct moving averages $\pi_t^{(j)}$ of length 2^j, i.e.,

$$\pi_t^{(j)} = \frac{1}{2^j} \sum_{p=0}^{2^j-1} x_{t-p}, \tag{10}$$

where $\pi_t^{(0)} \equiv x_t$. It is readily observed that these averages satisfy the recursion

$$\pi_t^{(j)} = \frac{\pi_t^{(j-1)} + \pi_{t-2^{j-1}}^{(j-1)}}{2}. \tag{11}$$

In essence, each element $\pi_t^{(j)}$ is an h-period moving average with $h = 2^j$.

Next, we denote by $\widehat{x}_t^{(j)}$ the difference between moving averages of lengths 2^{j-1} and 2^j, i.e.,

$$\widehat{x}_t^{(j)} = \pi_t^{(j-1)} - \pi_t^{(j)}. \tag{12}$$

Intuitively, the quantity $\widehat{x}_t^{(j)}$ captures fluctuations that survive the averaging over 2^{j-1} terms but disappear when the average involves 2^j terms, i.e., fluctuations with half-life in the interval $[2^{j-1}, 2^j)$. Consistent with this observation, the moving average, $\pi_t^{(j)}$, includes fluctuations whose half-life exceeds 2^j periods. From now on, we refer to the derived $\left\{\widehat{x}_t^{(j)}\right\}_{t\in\mathbb{Z}}$ as the filtered j-th component of the original time series $\{x_t\}_{t\in\mathbb{Z}}$. We note that these components are DHTs of the original process.

Since $\pi_t^{(0)} \equiv x_t$, by summing over j, it follows immediately from Eq. (12) that

$$x_t = \sum_{j=1}^{J} \widehat{x}_t^{(j)} + \pi_t^{(J)}, \tag{13}$$

for any $J \geq 1$. In other words, Eq. (13) decomposes the time series x_t into a sum of components with half-life belonging to a specific interval, plus a residual term $\pi_t^{(J)}$ which represents a long-run moving average. The larger J is, the closer the residual term $\pi_t^{(J)}$ is to a constant term. In practice, the choice of J is informed by empirical considerations dictated by the available sample length.

One may also write a convenient representation of the proposed filter using a suitable *orthonormal* operator. In order to illustrate how the operator works for $J = 2$, we first observe that, in this case, (11) yields

$$\pi_t^{(2)} = \frac{x_t + x_{t-1} + x_{t-2} + x_{t-3}}{4}. \tag{14}$$

Next, we substitute (11) into (12) and set $j = 1, 2$ to obtain

$$\widehat{x}_t^{(2)} = \frac{\pi_t^{(1)} - \pi_{t-2}^{(1)}}{2} = \frac{1}{2}\left(\frac{x_t + x_{t-1}}{2} - \frac{x_{t-2} + x_{t-3}}{2}\right), \tag{15}$$

$$\widehat{x}_t^{(1)} = \frac{\pi_t^{(0)} - \pi_{t-1}^{(0)}}{2} = \frac{x_t + x_{t-1}}{2}, \tag{16}$$

$$\widehat{x}_{t-2}^{(2)} = \frac{\pi_{t-2}^{(0)} - \pi_{t-3}^{(0)}}{2} = \frac{x_{t-2} + x_{t-3}}{2}. \tag{17}$$

This system of equations can now be expressed in matrix notation as

$$\begin{pmatrix} \pi_t^{(2)} \\ \widehat{x}_t^{(2)} \\ \widehat{x}_t^{(1)} \\ \widehat{x}_{t-2}^{(2)} \end{pmatrix} = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ \frac{1}{2} & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} x_t \\ x_{t-1} \\ x_{t-2} \\ x_{t-3} \end{pmatrix}. \tag{18}$$

Denoting by $\mathcal{T}^{(2)}$ the (4×4) matrix in (18), we notice that $\mathcal{T}^{(2)}$ is orthogonal, that is, $\Lambda^{(2)} \equiv \mathcal{T}^{(2)}(\mathcal{T}^{(2)})^\intercal$ is diagonal. Moreover, the diagonal elements of $\Lambda^{(2)}$ are non-vanishing so that $(\mathcal{T}^{(2)})^{-1} = (\mathcal{T}^{(2)})^\intercal(\Lambda^{(2)})^{-1}$ is well-defined. In essence, the matrix $\mathcal{T}^{(2)}$ permits construction of the filtered components by virtue of a simple matrix multiplication. Importantly, $\mathcal{T}^{(2)}$ leads to components sampled directly on the specific (sparse) support, $S_t^{(j)} = \{t - k2^j : k \in \mathbb{Z}\}$. We call sampling on the components' support, as opposed to calendar time, "decimated" sampling.

For an extension of this procedure to any $J \geq 2$, and a recursive algorithm for the construction of the matrix $\mathcal{T}^{(J)}$ associated to any arbitrary J value, we refer the reader to Mallat (1989). The matrix $\mathcal{T}^{(J)}$ is commonly known as the Haar Wavelet Transform (HWT).

The use of the HWT is a choice. Alternative nonparametric filters, like the Daubechies filter, could have been used instead without affecting the empirical results in the next section, but using a Haar filter is mathematically and logically convenient. Theorem 2 justifies this observation and formally shows that the Haar transform is effective in yielding an empirical evaluation of the beta representation in Eq. (6). We refer to Whitcher, Guttorp, and Percival (2000) for a related (auto-)covariance representation in the presence of overlapping observations.

Theorem 2 (Disaggregating β into spectral βs nonparametrically.) *Should a Haar transform be applied to the vector $\{x_t^1, x_t^2\}$ to decompose it into J decimated components, the* *resulting beta would conform with*

$$\widehat{\beta} = \frac{\widehat{\mathbb{C}}\left[x_t^1, x_t^2\right]}{\widehat{\mathbb{V}}\left[x_t^2\right]} = \sum_{j=1}^{J} \widehat{\upsilon}^{(j)} \widehat{\beta}^{(j)}, \tag{19}$$

where $\widehat{\beta}^{(j)} = \frac{\widehat{\mathbb{E}}\left[\widehat{x}_{k2^j}^{1,(j)},\, \widehat{x}_{k2^j}^{2,(j)}\right]}{\widehat{\mathbb{V}}\left[\widehat{x}_{k2^j}^{2,(j)}\right]}$ *and* $\widehat{\upsilon}^{(j)} = \frac{\widehat{\mathbb{V}}\left[\widehat{x}_{k2^j}^{2,(j)}\right]}{\widehat{\mathbb{V}}[x_t]}$.

Proof. *See Appendix* A.2.

We emphasize that the representation in Eq. (19) has two key features. First, the Haar transform delivers a beta expressed as a linear combination of betas defined with respect to *inner products* rather than with respect to *covariances*, thereby capturing (for all samples) the zero-mean nature of the Wold components in the extended Wold. Second, and more importantly, the cross-beta terms do not appear, thereby representing (for all samples) the uncorrelatedness, across frequencies, of the Wold components once more.

3 An Empirical Evaluation

The goal of this section is to implement the beta decomposition implied by the extended Wold representation in Eq. (2) and verify the identification potential of the two procedures presented in subsections 2.3 and 2.4.

To this end, we run market model-style regressions on two assets: a high book-to-market (value) portfolio and a low book-to-market (growth) portfolio.[3] The data are monthly and all returns are in excess of the risk-free rate. The results are:

$$R_{value,t} = \text{const}{+}1.149 \times R_{mkt,t} + \varepsilon_t, \quad R^2 = 0.90,$$
$$(\text{t-stat} = 31.69)$$
$$R_{growth,t} = \text{const}{+}0.973 \times R_{mkt,t} + \varepsilon_t, \quad R^2 = 0.90.$$
$$(\text{t-stat} = 38.19)$$

Since the volatility of the excess market return series in our sample is 19.06% per annum,

[3] We form quintile portfolios by sorting on the book-to-market variable as defined in Fama-French (1992). Similar results are obtained for terciles and deciles.

the beta estimates imply a covariance equal to $1.149 \times \left(19.06/\sqrt{12}\right)^2 = 34.784$ for value and $0.973 \times \left(19.06/\sqrt{12}\right)^2 = 29.456$ for growth. The regressions show the well-known result that these portfolios have a market beta around one.

The top two rows in Panels A and B of Table 1 focus on covariance decomposition. In Panel A, we obtain the Wold coefficients, the estimated residuals and, ultimately, the Wold components, as described in Subsection 2.3. Specifically, we fit a *companion* VAR(1) representation to the tri-variate series of market returns, value portfolio returns, and growth portfolio returns.[4] Using the Wold components, we then compute

$$\mathbb{C}\left(R_{mkt}^{(j)}, R_p^{(j)}\right),$$

where $p = \{value, growth\}$ and $j = 1, \dots, 6$. In Panel B, we instead use the nonparametric Haar-based filter described in Subsection 2.4.

The last column shows that, by summing the frequency-specific covariances, we are able to recover the overall covariances reported above. The small differences are due to the slightly different sample lengths caused by the lag structure in the VAR and the initialization of each component requiring a minimum of $2^j - 1$ observations (recall that the frequency-specific shocks at frequency j are a weighted sum of 2^j one-period shocks). The two methods agree on the covariances for frequencies $j \geq 3$, but they differ slightly on the values assigned to higher-frequency covariances (with $j < 3$). Because in the parametric procedure we impose a fairly simple structure, the small differences between parametric and nonparametric estimates are not surprising. If anything, it is interesting to note how well a simple parametric structure appears to capture first-order effects in the data.

The bottom rows in Panels A and B of Table 1 show the beta decomposition. We observe a widening of the dispersion in the betas associated with value and growth portfolios as we move from high frequencies (low j values) to low frequencies (high j values). We will return to this observation in subsequent sections. The last column shows an accurate reconstruction of the overall betas. As made explicit in Theorem 1, in order to obtain the overall beta, it is important to re-weight the frequency-specific betas according to

$$\sum_{j=1}^{J} w_p^{(j)} \beta_p^{(j)} \quad \text{with} \quad w_p^{(j)} = \frac{\mathbb{V}\left(R_{mkt}^{(j)}\right)}{\mathbb{V}\left(R_{mkt}\right)}.$$

[4] The lag length, p, of the original VAR is equal to 8. Results are robust to alternative choices around 8.

Panel A: Parametric

	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j>6$	$\sum_{j=1}^{J+1}\mathbb{C}(R_{mkt}^{(j)}, R_p^{(j)})$
Covariance decomposition								
Value	10.273	12.533	6.535	2.994	1.516	0.558	0.219	34.626
Growth	10.020	9.796	5.204	2.515	1.222	0.346	0.139	29.242

	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j>6$	$\sum_{j=1}^{J+1} w_p^{(j)}\beta_p^{(j)}$
Beta decomposition and reweighting								
Value	1.027	1.196	1.234	1.207	1.229	1.311	1.389	1.150
weight (rel. variance)	0.330	0.346	0.175	0.082	0.041	0.014	0.005	
Growth	1.003	0.950	0.961	1.000	0.987	0.813	0.883	0.965
weight (rel. variance)	0.330	0.340	0.179	0.083	0.041	0.014	0.005	

Panel B: Nonparametric

	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j>6$	$\sum_{j=1}^{J+1}\mathbb{C}(R_{mkt}^{(j)}, R_p^{(j)})$
Covariance decomposition								
Value	13.887	10.081	5.762	2.830	1.462	0.542	0.237	34.802
Growth	12.417	8.278	4.612	2.389	1.185	0.337	0.180	29.398

	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j>6$	$\sum_{j=1}^{J+1} w_p^{(j)}\beta_p^{(j)}$
Beta decomposition and reweighting								
Value	1.101	1.164	1.208	1.191	1.221	1.305	1.387	1.148
weight (rel. variance)	0.416	0.286	0.158	0.078	0.040	0.014	0.005	
Growth	0.984	0.956	0.967	1.005	0.989	0.812	0.882	0.969
weight (rel. variance)	0.416	0.286	0.158	0.078	0.040	0.014	0.005	

Table 1: All returns are in percent (i.e., raw covariances are multiplied by 10^4). Panel A displays a *parametric* decomposition of the covariances and betas (c.f., Subsection 2.3). Panel B displays a *nonparametric* decomposition of the covariances and betas (c.f., Subsection 2.4).

Finally, the differences that we observed in the estimated frequency-specific covariances between parametric and nonparametric estimates are much less pronounced when we turn to betas.

Next, we explore the orthogonality property of the proposed representation. To ease notation, we bundle together frequencies below one year and above one year. In practice, for each return series, we sum all of the components up to scale 4 (included) and dub this new component "the high-frequency component" (HF).[5] Analogously, for each return series, we sum all of the components higher than scale 4 and dub this new component "the low-frequency component" (LF). We then run the following simple regressions:

$$R_{p,t}^{LF} = \text{const} + \beta_p^{LF} \times R_{mkt,t}^{LF} + \varepsilon_t,$$
$$R_{p,t}^{HF} = \text{const} + \beta_p^{HF} \times R_{mkt,t}^{HF} + \varepsilon_t.$$

By the orthogonality of the components, the corresponding multiple regression, i.e.,

$$R_{p,t} = \text{const} + \beta_p^{HF} \times R_{mkt,t}^{HF} + \beta_p^{LF} \times R_{mkt,t}^{LF} + \varepsilon_t,$$

should deliver analogous beta estimates.[6] Table 2 displays the results.

Panel A: Parametric

	Simple Regression		Multiple Regression	
	β^{LF}	β^{HF}	β^{LF}	β^{HF}
	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Value	1.287	1.137	1.237	1.141
	(22.14)	(27.96)	(15.43)	(26.54)
Growth	0.937	0.986	0.874	0.986
	(18.03)	(34.38)	(11.24)	(33.48)

Panel B: Nonparametric

	Simple Regression		Multiple Regression	
	β^{LF}	β^{HF}	β^{LF}	β^{HF}
	(s.e.)	(s.e.)	(s.e.)	(s.e.)
Value	1.228	1.144	1.231	1.142
	(18.53)	(28.35)	(15.12)	(26.48)
Growth	0.947	0.985	0.891	0.983
	(20.50)	(34.39)	(11.78)	(32.84)

Table 2: Simple and multiple regression on high- and low-frequency betas.

Consistent with theoretical arguments regarding orthogonality of the components, we find that the multiple and simple regressions deliver similar estimates.

Once more, we observe a wider dispersion in the betas at low frequency with respect to high frequency, which is consistent with our previous results in Table 1, Panel B. Comparing Panels A and B in Table 2, we again see very similar results between the parametric and nonparametric methods.

[5]The cut-off $j = 4$ with monthly data corresponds to $2^4/12 = 1.\bar{3}$ year.

[6]We point out that, in the multiple regression, we do not decompose the regressand.

Our conclusions would not change if one were to choose a different cut-off j (except, of course, the spread between high-frequency betas and low-frequency betas would be more or less exacerbated). Similar results are also obtained if one estimates rolling regressions using a 10-year window from August of year $t-10$ to July of year t and averages the resulting betas across rolling samples. This is the approach we adopt in the out-of-sample evaluation of alternative portfolios in the next section.

4 Portfolio Selection

In this section, we present empirical evidence that the inclusion of a frequency dimension in the characterization of systematic risk may be used to select portfolios with lower out-of-sample variance relative to traditional multi-factor models.

4.1 The Optimization Problem(s)

In Markowitz's (1952) portfolio theory, given a target value $\tilde{\mu}$ for the expected portfolio return, the efficient portfolio weights, $\tilde{\mathbf{w}}$, are those that minimize the portfolio variance for all portfolios with expected return $\tilde{\mu}$. Mathematically, the optimization problem can be written as

$$\tilde{\mathbf{w}} \quad = \quad \arg\min_{\mathbf{w}} \mathbf{w}^{\intercal}\boldsymbol{\Sigma}\mathbf{w}, \tag{20}$$

subject to the constraints

$$\mathbf{w}^{\intercal}\boldsymbol{\mu} = \tilde{\mu} \quad \text{and} \quad \mathbf{w}^{\intercal}\mathbf{1} = 1, \tag{21}$$

where w_i is the portfolio weight on the i-th security, $\mu_i = \mathbb{E}[r_i]$ and $\Sigma_{i,j} = \mathbb{C}(r_i, r_j)$. A related optimzation problem minimizes variance in the absence of restrictions on the portfolio expected return (the global minimum variance problem). The weights may be constrained to be positive (long-*only* portfolios) or may be positive and negative (should short sales be allowed).

A key input in all optimization problems is, of course, the covariance matrix of the securities. In practice, the portfolio weights are optimized using a covariance matrix that has been estimated from historical data and are preserved until the next rebalancing. The out-of-sample performance of the covariance matrix can, therefore, be gauged by the variance of the optimal portfolio in the post-formation period. If the covariance estimator overfits the historical data, then out-of-sample performance can be poor, which is why imposing structure through a factor model can be beneficial. We will show that adding a frequency dimension to the factor(s) (and the resulting beta(s)) will increase the benefit of the assumed factor structures.

4.2 ***Spectral*** Factor Models

As is well-known, factor models of asset returns decompose the returns on a cross-section of assets into factor-related and asset-specific components.

Let $R_{i,t}$ denote the return on asset i in a universe of N stocks. Assuming M factors, vector $\beta_i = (\beta_{i,1}, \ldots, \beta_{i,M})$ gives asset i's sensitivities to the M factors, $f_t = (f_{1,t}, \ldots, f_{M,t})$. A factor decomposition of asset i's returns has the form

$$R_{i,t} = \alpha_i + \beta_i f_t^\intercal + \varepsilon_{i,t}. \tag{22}$$

It is commonly assumed that the asset-specific shocks, $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{N,t})$, are cross-sectionally uncorrelated so that $\mathbb{E}[\varepsilon_t^\intercal \varepsilon_t] = D$, where D is a diagonal matrix. Letting B be an $N \times M$-matrix of factor betas and V be the $M \times M$ covariance matrix of the factors, the covariance matrix of returns Σ_R can expressed as

$$\Sigma_R = BVB^\intercal + D. \tag{23}$$

Now, write a J-component spectral analogue to the model in Eq. (22), i.e.,

$$R_{i,t} = \alpha_i + \sum_{j=1}^{J} \beta_i^{(j)} (f_t^{(j)})^\intercal + \sum_{j=1}^{J} \varepsilon_{i,t}^{(j)}. \tag{24}$$

Since the Wold components are orthogonal to one another (c.f. Eq. (5)), we may write

$$\Sigma_{R,s} = \sum_{j=1}^{J} \Sigma^{(j)}, \tag{25}$$

$$\widehat{\Sigma}^{(j)} = B^{(j)} V^{(j)} {B^{(j)}}^\intercal + D^{(j)}. \tag{26}$$

Importantly, $\Sigma_R = \Sigma_{R,s}$ if $J = \infty$ and $\beta_i^{(j)} = \beta_i$ for all j. Thus, should J be large enough,[7] the difference between a traditional factor-based variance estimator, Σ_R, and our proposed variance estimator, $\Sigma_{R,s}$, would be induced by different levels of systematic risk across components. Differently put, the classical model in Eq. (22) can be viewed as a restricted version of the model we propose in Eq. (24), with equal levels of systematic risk across components.

We apply our spectral framework to the single-index market model of Sharpe (1963), the Fama-French three-factor model, and a five-factor model in which the factors are calculated using principal components analysis (PCA). Our interest is in comparing the out-of-sample performance of variance estimators based on traditional factor models and their spectral analogues.

4.3 The Data

The dataset and test criteria are similar to those in Ledoit and Wolf (2003). Stock return data are extracted from the University of Chicago's Center for Research in Securities Prices (CRSP) monthly database. Only U.S. common stocks traded on the New York Stock Exchange (NYSE) and the American Stock Exchange (AMEX) are included, which eliminates REIT's, ADR's, and other types of securities. Market returns are calculated as the equal-weighted portfolio returns given this universe of stocks. The size and book-to-market factors are extracted from Kenneth French's website at Dartmouth[8].

For $t = 1972$ to $t = 2016$, we use an in-sample period from August of year $t - 10$ to July of year t to form an estimate of the covariance matrix of stock returns. (We rebalance on the first trading day in August because AMEX stock return data become available in August 1962.) Future expected returns are estimated as the average realized return during the in-sample period.

Using these expected returns and the covariance matrix estimates, we form the global minimum variance portfolio (long only and long/short) and the minimum variance portfolio targeting a 20% expected return (with short sales allowed). The universe of stocks ranges between $N = 946$ and $N = 1302$.

4.4 Performance

Our measure of performance is the portfolio's out-of-sample standard deviation in the period from August 1972 to July 2016. For each factor model, we compute these standard deviations

[7] Technically, we are requiring the overall variance of the factor(s) to be close to the sum (for a finite J large enough) of the variances of the individual components of the factor(s), i.e., $V \approx \sum_{j=1}^{J} V^{(j)}$.

[8] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html

using non-spectral and spectral estimators of the covariance matrix. To assess robustness, we break the sample into periods: 1972 to 1994 (Table 3) and 1995 to 2016 (Table 4). The full period is in Tables 5 and 8. Consistent with the results in Section 3, the spectral models are implemented with $J = 6$.

Panel A in Tables 3 through 5 report the out-of-sample standard deviations of the minimum variance portfolios (with standard errors). In Panel B, we report the ratios of the out-of-sample standard deviations obtained from classical versus spectral models, along with one-sided p-values for tests of equal variances.

Allowing for changing systematic risk across frequencies translates into consistent improvements across models, optimization problems, and sample periods. As expected, the improvements are larger for factor models that are less accurately specified, i.e., they are larger with the market model than with the Fama-French three-factor model or PCA. Also, the improvements are larger when short sales are allowed. Even with the Fama-French model and PCA, however, the gains in terms of reduction in standard deviation can be between 5% and 10%, which is sizable in light of the out-of-sample nature of these approaches.

Panel A: Risk of minimum variance portfolios

	SD Global Min (Long Only)	SD Global Min	SD Min | E[R] = 20%
Market model	12.39 (0.53)	12.12 (0.52)	13.99 (0.60)
Spectral market model	12.24 (0.52)	11.67 (0.50)	13.22 (0.56)
Fama-French model	11.63 (0.50)	10.81 (0.46)	11.63 (0.50)
Spectral Fama-French model	11.60 (0.49)	10.44 (0.44)	11.28 (0.48)
Principal components	11.70 (0.50)	10.48 (0.45)	11.53 (0.49)
Spectral principal components	11.62 (0.50)	10.15 (0.43)	11.00 (0.47)

Panel B: F-test for equal variance

	SD Global Min (Long Only)	SD Global Min	SD Min | E[R] = 20%
Market model	1.02 (0.42)	1.08 (0.27)	1.12 (0.17)
Fama-French model	1.01 (0.48)	1.07 (0.28)	1.06 (0.31)
Principal components	1.01 (0.46)	1.06 (0.30)	1.10 (0.22)

Table 3: Panel A: Risk of minimum variance portfolios. We report standard deviations in percent and annualized through multiplication by $\sqrt{12}$. We also report the standard errors of these standard deviation estimates in parentheses. Panel B: Ratios of the non-spectral model variance to the spectral model variance are reported with their significance levels based on a one-sided F-test for equal variances. The time period is 1972 to 1994.

Panel A: Risk of minimum variance portfolios

	SD Global Min (Long Only)	SD Global Min	SD Min | E[R] = 20%
Market model	10.69 (0.48)	10.24 (0.46)	11.04 (0.49)
Spectral market model	10.69 (0.48)	9.77 (0.44)	10.45 (0.47)
Fama-French model	9.85 (0.44)	9.55 (0.43)	10.16 (0.45)
Spectral Fama-French model	9.85 (0.44)	9.41 (0.42)	9.88 (0.44)
Principal components	9.83 (0.44)	9.41 (0.42)	9.79 (0.44)
Spectral principal components	9.79 (0.44)	9.19 (0.41)	9.49 (0.42)

Panel B: F-test for equal variance

	SD Global Min (Long Only)	SD Global Min	SD Min | E[R] = 20%
Market model	1.00 (0.50)	1.10 (0.23)	1.12 (0.19)
Fama-French model	1.00 (0.51)	1.03 (0.41)	1.06 (0.33)
Principal components	1.01 (0.48)	1.05 (0.35)	1.07 (0.31)

Table 4: Panel A: Risk of minimum variance portfolios. We report standard deviations in percent and annualized through multiplication by $\sqrt{12}$. We also report the standard errors of these standard deviation estimates in parentheses. Panel B: Ratios of the non-spectral model variance to the spectral model variance are reported with their significance levels based on a one-sided F-test for equal variances. The time period is 1995 to 2016.

Panel A: Risk of minimum variance portfolios

	SD Global Min (Long Only)	SD Global Min	SD Min | E[R] = 20%
Market model	11.60 (0.36)	11.25 (0.35)	12.66 (0.39)
Spectral market model	11.52 (0.35)	10.81 (0.33)	11.97 (0.37)
Fama-French model	10.81 (0.33)	10.24 (0.32)	10.95 (0.34)
Spectral Fama-French model	10.80 (0.33)	9.97 (0.31)	10.63 (0.33)
Principal components	10.84 (0.33)	9.99 (0.31)	10.73 (0.33)
Spectral principal components	10.78 (0.33)	9.71 (0.30)	10.31 (0.32)

Panel B: F-test for equal variance

	SD Global Min (Long Only)	SD Global Min	SD Min | E[R] = 20%
Market model	1.01 (0.44)	1.08 (0.18)	1.12 (0.10)
Fama-French model	1.00 (0.49)	1.05 (0.27)	1.06 (0.25)
Principal components	1.01 (0.45)	1.06 (0.26)	1.08 (0.18)

Table 5: Panel A: Risk of minimum variance portfolios. We report standard deviations in percent and annualized through multiplication by $\sqrt{12}$. We also report the standard errors of these standard deviation estimates in parentheses. Panel B: Ratios of the non-spectral model variance to the spectral model variance are reported with their significance levels based on a one-sided F-test for equal variances.The time period is 1972 to 2016.

5 The Economics of Spectral Betas

We now return to the logic behind the empirical findings in the previous section. In order to do so, we report in Fig. 2 the average frequency-specific market betas for portfolios that are sorted by size from 1972 to 2016 and book-to-market from 1980 to 2016.[9] We also report the estimated frequency-specific alphas of Sharpe's (1963) single index market model using frequency-specific market betas in place of the overall market beta. The results complement those reported in Section 3.

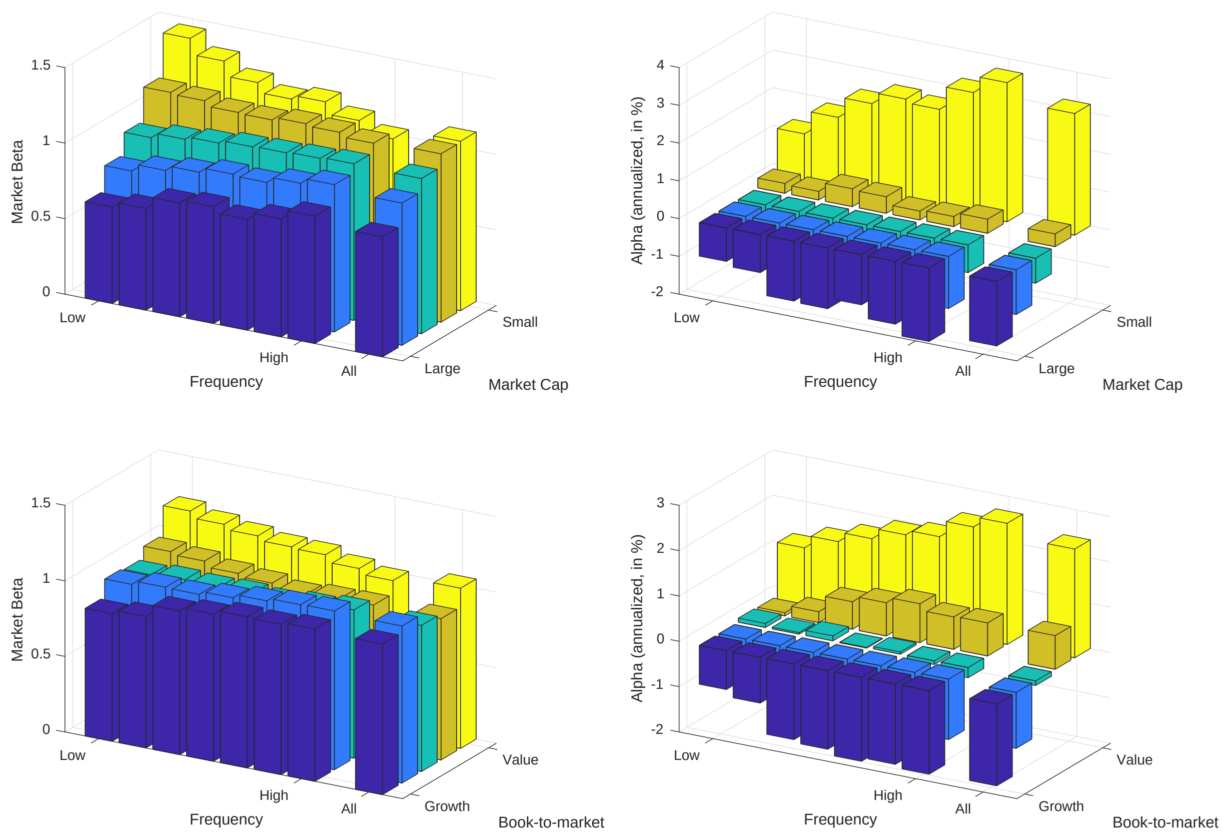

Figure 2: Spectral market model betas and corresponding alphas. Average frequency-specific market betas for portfolios that are sorted by size from $t = 1972$ to 2016 and book-to-market from $t = 1980$ to 2016. The alphas of Sharpe's (1963) single index market model using spectral market betas in place of the overall market beta are also reported.

The spectral betas vary not only across the size and value dimensions, but also along the frequency dimension. Importantly, the change across frequencies may be larger than the change across styles. Also, the spectral betas either decline (for large or growth companies) or increase (for small or value companies) almost monotonically. Small and value stocks

[9]Book value is extracted from Compustat and only becomes available to match our CRSP data around 1980. The universe of stocks for this book-to-market analysis varies from $N = 703$ to $N = 885$.

are particularly sensitive to market fluctuations at low frequencies, while large and growth stocks are the opposite. Because the betas at the lowest frequencies capture fluctuations between $2^6 = 64$ months and $2^7 = 128$ months, this finding suggests that small and value stocks are more affected by dynamics around the long end of the business cycle.

5.1 A frequency-specific beta CAPM

The right panels in Fig. 2 reports the pricing errors from a CAPM model based on spectral betas. Specifically, given a frequency band j, we compute

$$\alpha_i^{(j)} = \mathbb{E}[R_{i,t} - R_f] - \beta_i^{(j)}\mathbb{E}[R_{mkt,t} - R_f], \tag{27}$$

where we have restricted the price of risk across frequencies to be equal to the unconditional market risk premium. We call this model a frequency-specific beta CAPM.[10] This model is useful to explore whether components of market risk with heterogeneous periodicities are important drivers of risk premia.

The estimated alphas in Fig. 2 decrease as we extend the horizon. This result is the flip side of the finding discussed in the previous section: the risk exposures (i.e., the betas) align better with expected returns as we move from high and low frequencies.

These results are suggestive of the fact that a *single-index* model performs more convincingly when the search for priced systematic risk occurs at low frequencies. To further explore this venue, we turn to the widely analyzed, and yet somewhat challenging, cross-section of 25 Fama-French portfolios double sorted on size and book-to-market. Fig. 3 shows the behavior of the corresponding spectral betas across alternative scales. Fig. 4 displays the pricing errors from a CAPM model which uses spectral betas to capture the frequency-specific nature of risk, c.f. Eq. (27).[11]

Focusing on the risk exposures, we observe that the betas corresponding to lower frequency components of the market process ($j = 4, \ldots, 7$) display improved alignment in the value space without losing their ability to capture cross-sectional variability in risk premia across size quintiles. If anything, we also observe an improvement in the size space as a function of the horizon length. As one would expect, this enhanced alignment has important

[10] One can also explore a CAPM model with frequency-specific betas *and* frequency-specific prices of risk. Such a specification, however, would require thinking carefully about notions of frequency-specific risk-aversion. We hope the findings in this paper will encourage investigations in this direction.

[11] The betas and alphas are estimated on a rolling basis (and then averaged) to parallel the portfolio optimization problem. A cross-sectional regression with betas estimated over the full sample leaves unchanged our conclusions.

implications for cross-sectional pricing. Indeed, the pricing errors are substantially reduced as we move from high to low frequencies.

Some observations are in order. First, we re-emphasize that, given the structure of the model (which uses just one j-specific beta at a time), a decrease in alphas as the time horizon increases suggests that low-frequency betas yield superior fit. Our results are, therefore, consistent with — and re-interpret from a different perspective — the recent work that has highlighted the role played by hard-to-detect persistent components in risk factors – like consumption – for asset prices (e.g., Alvarez and Jermann (2005), Bansal and Yaron (2004), and Hansen (2008)). Risk, in our framework, is driven by slow-moving market components.

Second, we have used suitably-defined low-frequency betas to explain *short-term* returns. In other words, our emphasis is on the importance of low frequencies in the assessment of risk at all frequencies (the shortest one, in particular), rather than on the evaluation of a classical model at different horizons.

Finally, it is instructive to benchmark the proposed specifications against a 3-factor model based on HML, SMB and the market portfolio. Such a model attains a RMSE and a MAPE equal to 1.17% and 1.76% per year, respectively. A one-factor model based on long-term ($j = 7$) spectral betas improves upon the 3-factor Fama-French model and, instead, achieves a RMSE of 0.958% and a MAPE of 1.39%. We argue that explicit allowance for a separation between fluctuations with heterogeneous durations in market risk — and, more generally, in any risk factor — is important for evaluating cross-sectional pricing. We view this finding as being particularly stimulating in light of the recent literature investigating the number of factors needed to explain the cross-section of stock returns (*inter alia*, DeMiguel, Martin-Utrera, Nogales, and Uppal (2017), Harvey, Liu, and Zhu (2016), Feng, Giglio, and Xiu (2017), Freyberger, Neuhierl, and Weber (2017), Kelly, Pruitt, and Su (2018), Kozak, Nagel, and Santosh (2017)). Our evidence suggests that computing exposures to risk fluctuations with heterogeneous duration may provide a revealing new tool to achieve satisfactory dimension reduction while preserving the economic appeal of well-understood factors (like the market).

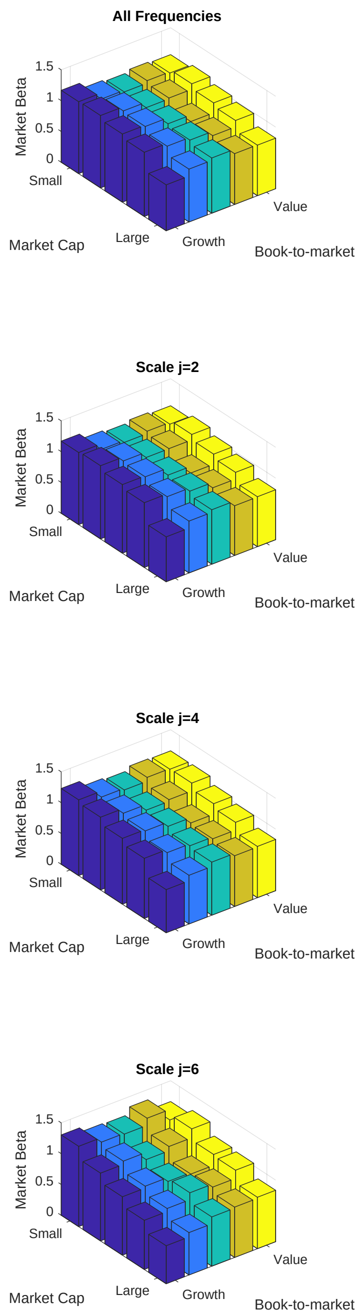

Figure 3: Spectral market betas. Average spectral market betas for the 25 portfolios double-sorted by size and book-to-market. The sample period is 1980 to 2016.

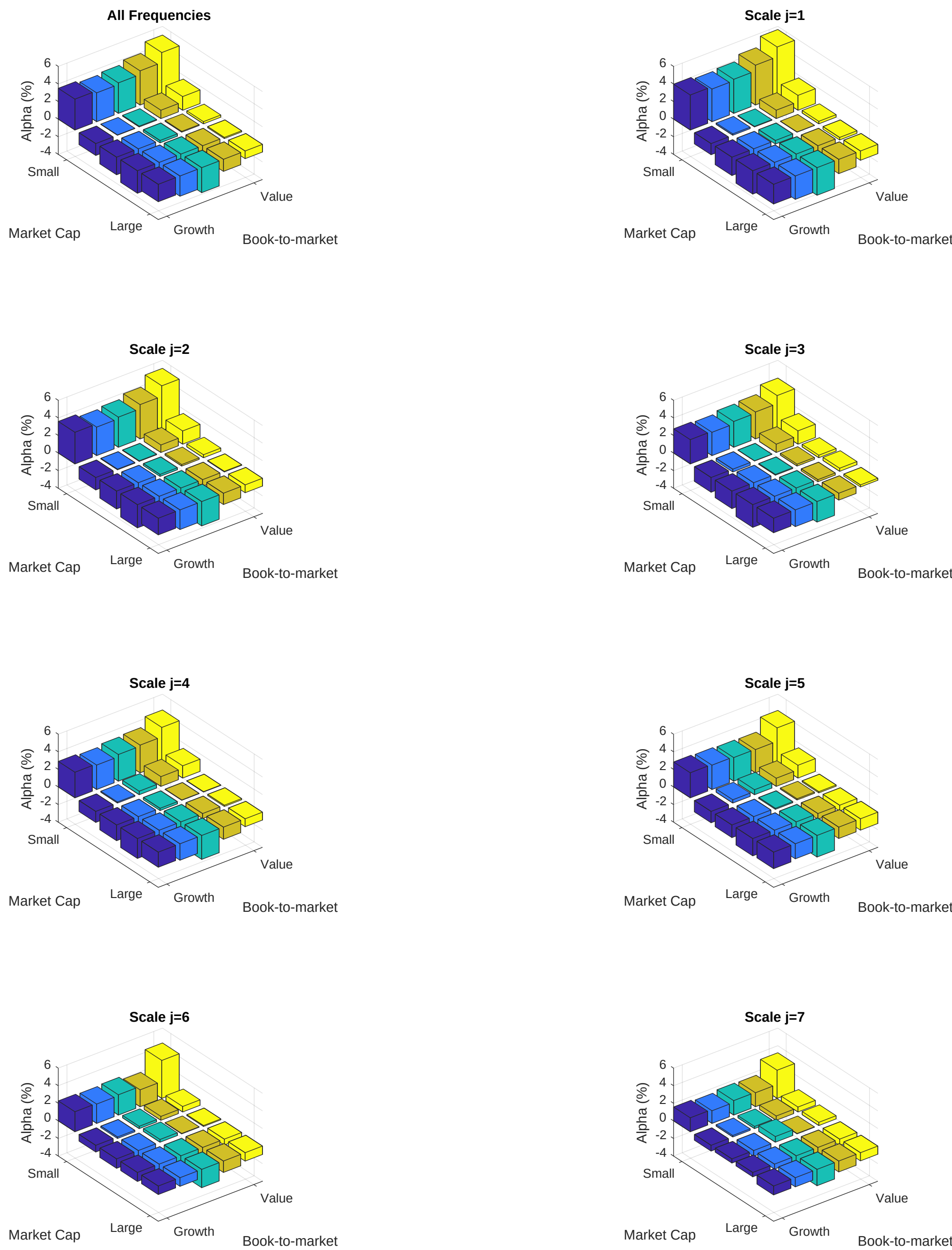

Figure 4: Alphas from spectral market betas. Average alphas for the 25 portfolios double-sorted by size and book-to-market. The sample period is 1980 to 2016.

6 Conclusion

Using a suitable Wold representation, we introduce the notion of spectral beta, i.e., the beta between frequency-specific components of a multivariate process. We show how the traditional beta can naturally be expressed as a linear combination of spectral or frequency-specific betas. Importantly, our proposed beta representation does not involve cross-covariances or cross-betas between frequency-specific components, thereby permitting all frequency-specific information to be contained exclusively in the frequency-specific betas. The representation is in the time domain, something which facilitates economic interpretation and applicability.

The spectral betas allow us to characterize heterogeneity in systematic risk across frequencies, thereby providing an additional key dimension to portfolio allocation, risk management, and asset pricing. Using NYSE and AMEX stock returns from 1972 to 2016, we find that spectral factor models lead to the selection of portfolios with significantly lower out-of-sample variance relative to their traditional analogues. We have also shown that a spectral CAPM model in which systematic risk is measured by way of a *single* low-frequency spectral beta, and the price of risk is set as being equal to the unconditional market risk premium, may achieve dimension reduction as compared to popular multi-factor models while preserving the centrality of the market as a key source of systematic risk.

If the quantities of risk have a frequency component to them, do the prices of risk also have a frequency component to them? While our methodological focus is on spectral betas given constant prices of risk across frequencies, we provide methods to operationalize heterogeneous pricing across frequencies (without necessarily implying it). In fact, we view the methods as providing an empirical laboratory to move past a classical restriction in the asset pricing literature, that of equal prices of risk across frequencies, and inform structural modeling, if the restriction is violated. We hope the laboratory will stimulate future investigations in this direction.

A Appendix

A.1 Corollary to Theorem 1.

Assume $\mathbf{x} = \{(x_t^1, x_t^2)^\intercal\}_{t\in\mathbb{Z}}$ satisfies (1). The resulting beta would conform with

$$\beta = \frac{\mathbb{C}\left[x_t^1, x_t^2\right]}{\mathbb{V}\left[x_t^2\right]} = \sum_{j=1}^{\infty} v^{(j)} \beta^{(j)},$$

where

$$\begin{aligned}
\beta^{(j)} \quad &= \quad \frac{\sum_{k=0}^{\infty} \sigma^2_{\varepsilon^1} \left(\sum_{i=0}^{2^{j-1}-1} \alpha^1_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha^1_{k2^j+2^{j-1}+i}\right)\left(\sum_{i=0}^{2^{j-1}-1} \alpha^3_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha^3_{k2^j+2^{j-1}+i}\right)}{\mathbb{V}\left[x_t^{2,(j)}\right]} \\
&+ \quad \frac{\sum_{k=0}^{\infty} \sigma^2_{\varepsilon^2} \left(\sum_{i=0}^{2^{j-1}-1} \alpha^2_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha^2_{k2^j+2^{j-1}+i}\right)\left(\sum_{i=0}^{2^{j-1}-1} \alpha^4_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha^4_{k2^j+2^{j-1}+i}\right)}{\mathbb{V}\left[x_t^{2,(j)}\right]} \\
&+ \quad \frac{\sum_{k=0}^{\infty} \sigma_{\varepsilon^{1,2}} \left(\sum_{i=0}^{2^{j-1}-1} \alpha^1_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha^1_{k2^j+2^{j-1}+i}\right)\left(\sum_{i=0}^{2^{j-1}-1} \alpha^4_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha^4_{k2^j+2^{j-1}+i}\right)}{\mathbb{V}\left[x_t^{2,(j)}\right]} \\
&+ \quad \frac{\sum_{k=0}^{\infty} \sigma_{\varepsilon^{1,2}} \left(\sum_{i=0}^{2^{j-1}-1} \alpha^2_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha^2_{k2^j+2^{j-1}+i}\right)\left(\sum_{i=0}^{2^{j-1}-1} \alpha^3_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha^3_{k2^j+2^{j-1}+i}\right)}{\mathbb{V}\left[x_t^{2,(j)}\right]}
\end{aligned} \tag{A.1}$$

with

$$\begin{aligned}
\mathbb{V}\left[x_t^{2,(j)}\right] \quad &= \quad \sum_{k=0}^{\infty} \sigma^2_{\varepsilon^1} \left(\sum_{i=0}^{2^{j-1}-1} \alpha^3_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha^3_{k2^j+2^{j-1}+i}\right)^2 \\
&+ \quad \sum_{k=0}^{\infty} \sigma^2_{\varepsilon^2} \left(\sum_{i=0}^{2^{j-1}-1} \alpha^4_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha^4_{k2^j+2^{j-1}+i}\right)^2 \\
&+ \quad \sum_{k=0}^{\infty} 2\sigma_{\varepsilon^{1,2}} \left(\sum_{i=0}^{2^{j-1}-1} \alpha^3_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha^3_{k2^j+2^{j-1}+i}\right)\left(\sum_{i=0}^{2^{j-1}-1} \alpha^4_{k2^j+i} - \sum_{i=0}^{2^{j-1}-1} \alpha^4_{k2^j+2^{j-1}+i}\right)
\end{aligned}$$

and

$$v^{(j)} = \frac{\mathbb{V}\left[x_t^{2,(j)}\right]}{\sum_{k=0}^{\infty}(\sigma^2_{\varepsilon^1}(\alpha^3_k)^2 + \sigma^2_{\varepsilon^2}(\alpha^4_k)^2 + 2\sigma^2_{\varepsilon^{1,2}}\alpha^3\alpha^4)}. \tag{A.2}$$

Proof. Immediate, given the representation in (2) and its properties.

A.2 Proof of Theorem 2.

We assume availability of $T = 2^J$ observations. This is, of course, always true since observations can be trimmed. We denote by $x_T^{1,(J)}$ and $x_T^{2,(J)}$ the T-vectors collecting the observations of the series $\{x_t^1\}$ and $\{x_t^2\}$, respectively:

$$
\begin{aligned}
x_T^{1,(J)} &= \left[x_T^1,\ x_{T-1}^1,\ \ldots,\ x_1^1\right]^\intercal ,\\
x_T^{2,(J)} &= \left[x_T^2,\ x_{T-1}^2,\ \ldots,\ x_1^2\right]^\intercal .
\end{aligned}
$$

Define, also, the vectors of *decimated* components $\mathbf{x}^{\mathbf{1},(\mathbf{j})} = \left[x_{2^j}^{1,(j)}, x_{2\times 2^j}^{1,(j)}, x_{3\times 2^j}^{1,(j)}, ...\right]^\intercal$ and $\mathbf{x}^{\mathbf{2},(\mathbf{j})} = \left[x_{2^j}^{2,(j)}, x_{2\times 2^j}^{2,(j)}, x_{3\times 2^j}^{2,(j)}, ...\right]^\intercal$. Now, write the empirical covariance of the two series as

$$
\begin{aligned}
\frac{\sum_{t=1}^T x_t^1 x_t^2}{T} - \frac{\sum_{t=1}^T x_t^1}{T}\frac{\sum_{t=1}^T x_t^2}{T} &= \frac{\left(x_T^{1,(J)}\right)^\intercal \left(x_T^{2,(J)}\right)}{T} - \frac{\sum_{t=1}^T x_t^1}{T}\frac{\sum_{t=1}^T x_t^2}{T}\\
&= \frac{\left(x_T^{1,(J)}\right)^\intercal (\mathcal{T}^{(J)})^\intercal (\Lambda^J)^{-1} (\mathcal{T}^{(J)}) \left(x_T^{2,(J)}\right)}{T} - \frac{\sum_{t=1}^T x_t^1}{T}\frac{\sum_{t=1}^T x_t^2}{T}\\
&= \frac{\left(\left(\Lambda^{(J)}\right)^{-1/2} \mathcal{T}^{(J)} x_T^{1,(J)}\right)^\intercal \left(\left(\Lambda^{(J)}\right)^{-1/2} \mathcal{T}^{(J)} x_T^{2,(J)}\right)}{T} - \frac{\sum_{t=1}^T x_t^1}{T}\frac{\sum_{t=1}^T x_t^2}{T}\\
&= \frac{\sum_{j=1}^J 2^j (\mathbf{x}^{\mathbf{1},(\mathbf{j})})^\intercal \mathbf{x}^{\mathbf{2},(\mathbf{j})}}{T} + 2^J \frac{\pi_{x^1}^{(J)} \pi_{x^2}^{(J)}}{T} - \frac{\sum_{t=1}^T x_t^1}{T}\frac{\sum_{t=1}^T x_t^2}{T}\\
&= \sum_{j=1}^J \frac{(\mathbf{x}^{\mathbf{1},(\mathbf{j})})^\intercal \mathbf{x}^{\mathbf{2},(\mathbf{j})}}{\frac{T}{2^j}} ,
\end{aligned}
$$

where, in the second equality, we exploit the fact that the matrix $(\mathcal{T}^{(J)})^\intercal (\Lambda^J)^{-1} (\mathcal{T}^{(J)})$ is the identity matrix, in the fourth equality we use the fact that the diagonal elements of the matrix $\Lambda^{(J)} \equiv \mathcal{T}^{(J)} \left(\mathcal{T}^{(J)}\right)^T$ are $\lambda_1 = \lambda_2 = 1/2^J$ and $\lambda_k = 1/2^{J-j+1}$ for $k = 2^{j-1}+1, \ldots, 2^j$, $j = 2, \ldots, J$, and in the last row we use the definition of the scaling components, $\pi_{x^1}^{(J)}$ and $\pi_{x^2}^{(J)}$. Next, observe that the components $\mathbf{x}^{\mathbf{1},(\mathbf{j})}$ and $\mathbf{x}^{\mathbf{2},(\mathbf{j})}$ at scale j have exactly $\frac{T}{2^j}$ observations due to decimation. □

References

Alvarez, F., and U. J. Jermann (2005): "Using Asset Prices to Measure the Persistence of the Marginal Utility of Wealth," *Econometrica*, 73(6), 1977–2016.

Bandi, F. M., B. Perron, A. Tamoni, and C. Tebaldi (2017): "The scale of predictability," *Journal of Econometrics, forthcoming.*

Bansal, R., and A. Yaron (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59(4), 1481–1509.

Berkowitz, J. (2001): "Generalized spectral estimation of the consumption-based asset pricing model," *Journal of Econometrics*, 104(2), 269–288.

Beveridge, S., and C. R. Nelson (1981): "A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'," *Journal of Monetary Economics*, 7(2), 151–174.

Cogley, T. (2001): "A Frequency Decomposition of Approximation Errors in Stochastic Discount Factor Models," *International Economic Review*, 42(2), 473–503.

DeMiguel, V., A. Martin-Utrera, F. J. Nogales, and R. Uppal (2017): "A Portfolio Perspective on the Multitude of Firm Characteristics," Working paper, LBS.

Dew-Becker, I., and S. Giglio (2016): "Asset Pricing in the Frequency Domain: Theory and Empirics," *The Review of Financial Studies*, 29(8), 2029–2068.

Fama, E. F., and K. R. French (1992): "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47(2), 427–465.

Feng, G., S. Giglio, and D. Xiu (2017): "Taming the Factor Zoo," Fama-Miller Working Paper 17-04, Chicago Booth Research Paper.

Freyberger, J., A. Neuhierl, and M. Weber (2017): "Dissecting Characteristics Nonparametrically," CESifo Working Paper Series 6391, CESifo Group Munich.

Hannan, E. J. (1963a): *Regression for Time Series.* in Time Series Analysis, ed. by M. Rosenblatt. New York: Wiley.

——— (1963b): "Regression for Time Series with Errors of Measurement," *Biometrika*, 50, pp. 293–302.

Hansen, L. P. (2008): "Modeling the Long Run: Valuation in Dynamic Stochastic Economies," NBER Working Papers 14243, National Bureau of Economic Research, Inc.

Harvey, C. R., Y. Liu, and H. Zhu (2016): "...and the Cross-Section of Expected Returns," *The Review of Financial Studies*, 29(1), 5–68.

Kelly, B. T., S. Pruitt, and Y. Su (2018): "Characteristics Are Covariances: A Unified Model of Risk and Return," Fama-Miller Working Paper 17-04, SSRN.

Kozak, S., S. Nagel, and S. Santosh (2017): “Interpreting Factor Models,” *Journal of Finance - Forthcoming*.

Ledoit, O., and M. Wolf (2003): “Improved estimation of the covariance matrix of stock returns with an application to portfolio selection,” *Journal of Empirical Finance*, 10(5), 603–621.

Lintner, J. (1965): “The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets,” *The Review of Economics and Statistics*, 47(1), 13–37.

Lutkepohl, H. (2005): *New introduction to multiple time series analysis*. Springer.

Mallat, S. G. (1989): “A Theory for Multiresolution Signal Decomposition: The Wavelet Representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, 11, 674–693.

Markowitz, H. (1952): “Portfolio Selection,” *Journal of Finance*, 7(1), 77–91.

Ortu, F., F. Severino, A. Tamoni, and C. Tebaldi (2017): “A persistence-based Wold-type decomposition for stationary time series,” *SSRN eLibrary*.

Sharpe, W. F. (1963): “A Simplified Model for Portfolio Analysis,” *Management Science*, 9(2), 277–293.

——— (1964): “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk,” *Journal of Finance*, 19(3), 425–442.

Whitcher, B., P. Guttorp, and D. B. Percival (2000): “Wavelet analysis of covariance with application to atmospheric time series,” *Journal of Geophysical Research: Atmospheres*, 105(D11), 14941–14962.

Wold, H. (1938): *A Study in the Analysis of Stationary Time Series*. Uppsala, Sweden: Almqvist and Wiksell.