Principal Component Analysis of High Frequency Data*

Yacine Aït-Sahalia†  
Department of Economics  
Princeton University and NBER

Dacheng Xiu‡  
Booth School of Business  
University of Chicago

This Version: October 7, 2016

Abstract

We develop the necessary methodology to conduct principal component analysis at high frequency. We construct estimators of realized eigenvalues, eigenvectors, and principal components and provide the asymptotic distribution of these estimators. Empirically, we study the high frequency covariance structure of the constituents of the S&P 100 Index using as little as one week of high frequency data at a time, and examines whether it is compatible with the evidence accumulated over decades of lower frequency returns. We find a surprising consistency between the low and high frequency structures. During the recent financial crisis, the first principal component becomes increasingly dominant, explaining up to 60% of the variation on its own, while the second principal component drives the common variation of financial sector stocks.

Keywords: Itô Semimartingale, High Frequency, Spectral Function, Eigenvalue, Eigenvector, Principal Components, Three Factor Model.

JEL Codes: C13, C14, C55, C58.

*We thank Hristo Sendov for valuable discussions on eigenvalues and spectral functions. In addition, we benefited much from comments by Torben Andersen, Tim Bollerslev, Oleg Bondarenko, Marine Carrasco, Gary Chamberlain, Kirill Evdokimov, Jianqing Fan, Christian Hansen, Jerry Hausman, Jean Jacod, Ilze Kalnina, Jia Li, Yingying Li, Oliver Linton, Nour Meddahi, Per Mykland, Ulrich Müller, Andrew Patton, Eric Renault, Jeffrey Russell, Neil Shephard, George Tauchen, Viktor Todorov, Ruey Tsay, and Xinghua Zheng, as well as seminar and conference participants at Brown University, CEMFI, Duke University, Harvard University, MIT, Northwestern University, Peking University, Princeton University, Singapore Management University, Stanford University, University of Amsterdam, University of Chicago, University of Illinois at Chicago, University of Tokyo, the 2015 North American Winter Meeting of the Econometric Society, the CEME Young Econometricians Workshop at Cornell, the NBER-NSF Time Series Conference in Vienna, the 10th International Symposium on Econometric Theory and Applications, the 7th Annual SoFiE Conference, the 2015 Financial Econometrics Conference in Toulouse and the 6th French Econometrics Conference.

†Address: 26 Prospect Avenue, Princeton, NJ 08540, USA. E-mail address: yacine@princeton.edu.
‡Address: 5807 S Woodlawn Avenue, Chicago, IL 60637, USA. E-mail address: dacheng.xiu@chicagobooth.edu.

Xiu gratefully acknowledges financial support from the IBM Faculty Scholar Fund at the University of Chicago Booth School of Business.
1 Introduction

Principal component analysis (PCA) provide information about any latent common structure that might exist in a dataset. As a result, it is one of the most popular techniques in multivariate statistics, especially when analyzing large datasets (see, e.g., Hastie, Tibshirani, and Friedman (2009).) The central idea of PCA is to identify a small number of common or principal components which effectively summarize a large part of the variation of the data, and serve to reduce the dimensionality of the problem and achieve parsimony.

Classical PCA originated in Pearson (1901) and Hotelling (1933) and is widely used in macroeconomics and finance, often as a precursor or in addition to factor analysis. Just like a non-parametric regression is useful to get an idea of what parametric functional form to use later on, PCA is useful to find out whether a factor model is even a good idea for a given dataset, how many factors to employ, etc. Examining a PCA scree plot is a natural preliminary step before employing a factor model. PCA has indeed been used extensively at low frequency to specify factor models and to determine the number of factors (see Bai (2003).) Among applications, Litterman and Scheinkman (1991) employ PCA to document a three-factor structure of the term structure of bond yields, Stock and Watson (1999) and Stock and Watson (2002a) to construct factors for forecasting, Stock and Watson (2002b) to develop economic activity and inflation indices, while Egloff, Leippold, and Wu (2010) suggest a two-factor model for volatility based on PCA. A sentiment measure, reflecting the optimistic or pessimistic view of investors, was created by Baker and Wurgler (2006) using the first principal component of a number of sentiment proxies, while policy uncertainty index was created by Baker, Bloom, and Davis (2013). PCA can also be employed to analyze factor structures that are only approximate, as in Chamberlain and Rothschild (1983) compared to Ross (1976), including estimating the number of approximate factors (see Connor and Korajczyk (1993), Bai and Ng (2002), Amengual and Watson (2007), Onatski (2010) and Kapetanios (2010).)

The estimation of the eigenvalues of the sample covariance matrix is the key step towards PCA. Classical low frequency asymptotic results, starting with Anderson (1958) and Anderson (1963), show that eigenvalues of the sample covariance matrix are consistent and asymptotically normal estimators of the population eigenvalues, at least when the data follow a multivariate normal distribution. Even under normality, the asymptotic distribution becomes rather involved when repeated eigenvalues are present. Waternaux (1976) showed that a similar central limit result holds for simple eigenvalues as long as the distribution of the data has finite fourth moment, while Tyler (1981) and Eaton and Tyler (1991) obtain the asymptotic distribution of eigenvectors and eigenvalues under more general assumptions; see also discussions in the books by Jolliffe (2002) and Jackson (2003).

The classical PCA approach to statistical inference of eigenvalues suffers from three main
drawbacks. First is the curse of dimensionality. For instance, it is well known that even the largest
eigenvalue is no longer consistently estimated when the cross-sectional dimension grows at the
same rate as the sample size along the time domain. Second, the asymptotic theory is essentially
dependent on frequency domain analysis under stationarity (and often additional) assumptions,
see, e.g., Brillinger (2001). Third, the principal components are linear combinations of the data,
which fail to capture potentially nonlinear patterns therein.

These three drawbacks create difficulties when PCA is employed on asset returns data. A
typical portfolio may consist of dozens of stocks. For instance, a portfolio with 30 stocks has 465
parameters in their covariance matrix, if no additional structure is imposed, while one with 100
stocks contain 5,050 parameters. As a result, years of time series data are required for estimation,
raising issues of survivorship bias, potential non-stationarity and parameter constancy. Moreover,
asset returns are known to exhibit time-varying volatility and heavy tails, leading to deviations from
the assumptions required by the classical PCA asymptotic theory. In addition, prices of derivatives,
e.g., options, are often nonlinear functions of the underlying asset’s price and volatility. Ruling out
nonlinear combinations disconnects their principal components from the underlying factors that
drive derivative returns.

These issues motivate the development in this paper of the tools necessary to conduct PCA for
continuous-time stochastic processes using high frequency data and high frequency or “in-fill” as-
symptotics, whereby the number of observations grows within a fixed time window. This approach
not only adds the important method of PCA to the high frequency toolkit, but it also addresses
the three drawbacks mentioned above. First, the large amount of intraday time series data vastly
improve the time series vs. cross-sectional dimensionality trade-off. And these are not redundant
observations for the problem at hand. Unlike for expected returns, it is well established theoreti-
cally that increasing the sampling frequency improves variance and covariance estimation, at least
until market microstructure concerns start biting. Yet market microstructure noise is not a serious
concern at the one minute sampling frequency we will employ empirically, for liquid stocks which
typically trade on infra-second time scales. As a result of the large increase in the time series
dimension, it is plausible to expect high frequency asymptotic analysis with the cross-sectional
dimension fixed to serve as accurate approximations.\(^1\) In fact, we will show both in simulations

\(^1\) At low frequency, in the case of large dimensional data, Geman (1980) and Bai, Silverstein, and Yin (1988)
investigated the strong law of the largest eigenvalue, when the cross-sectional dimension of the data grows at the
same rate as the sample size. Johnstone (2001) further analyzed the distribution of the largest eigenvalue using
random matrix theory. These analyses confirmed that the largest eigenvalue is no longer consistently estimated.
Moreover, Johnstone and Lu (2009) prove that, for a single factor model, the estimated eigenvector corresponding
to the largest eigenvalue is also inconsistent unless the sample size grows at a rate faster than the one at which
the cross-sectional dimension increases. To resolve the inconsistency of the PCA in this setting, different estimators
of the covariance matrix have been proposed, using banding (Bickel and Levina (2008b)), tapering (Cai and Zhou
(2012)), and thresholding (Bickel and Levina (2008a)) methods, or imposing additional assumptions such as sparsity.
However, the sparsity requirement of the covariance matrix does not hold empirically or a large cross-section of stock
returns (see, e.g., Fan, Furger, and Xiu (2016).) Alternatively, there is a new literature on methods that “sparsify”
and empirically that PCA works quite well at high frequency for a portfolio of 100 stocks using as little as one week of one-minute returns. Second, the high frequency asymptotic framework enables nonparametric analysis of general stochastic processes, namely Itô semimartingales, thereby allowing strong dependence, non-stationarity, and heteroscedasticity due to stochastic volatility and jumps, and freeing the analysis from the strong parametric assumptions that are required in a low frequency setting. In effect, the existence of population moments becomes irrelevant and PCA becomes applicable at high frequency for general Itô semimartingales. Third, our principal components are built upon instantaneous (or locally) linear combinations of the stochastic processes, thereby capturing general nonlinear relationships by virtue of Itô’s lemma.

Within a continuous-time framework, we first develop the concept of “realized” or high-frequency PCA for data sampled from a stochastic process within a fixed time window. Realized PCA depends on realizations of eigenvalues, which are stochastic processes, evolving over time. Our implementation of realized PCA is designed to extend classical PCA from its low frequency setting into the high frequency one as seamlessly as possible. Therefore, we start by estimating realized eigenvalues from the realized covariance matrix. One technical challenge we must overcome is the fact that when an eigenvalue is simple, it is a smooth function of the instantaneous covariance matrix. On the other hand, when it comes to a repeated eigenvalue, this function is not differentiable, which hinders statistical inference, as the asymptotic theory requires at least second-order differentiability. To tackle the issue of non-smoothness of repeated eigenvalues, we propose an estimator constructed by averaging all repeated eigenvalues. Using the theory of spectral functions, we show how to obtain differentiability of the proposed estimators, and the required derivatives. We therefore construct an estimator of realized spectral functions, and develop the high frequency asymptotic theory for this estimator. Estimation of eigenvalues, principal components and eigenvectors then arise as special cases of this estimator, by selecting a specific spectral function. Aggregating local estimates results in an asymptotic bias, due to the nonlinearity of the spectral functions. The bias is fortunately of higher order, which enables us to construct a bias-corrected estimator, and show that the latter is consistent and achieves stable convergence in law to a mixed normal distribution.

As far as asymptotic theory is concerned, our estimators are constructed by aggregating functions of instantaneous covariance estimates. Other examples of this general type arise in the high frequency context. Jacod and Rosenbaum (2013) analyze such estimators with the objective of estimating the integrated quarticity. Li, Todorov, and Tauchen (2016) estimate volatility functional dependencies, while Kalnina and Xiu (2016) estimate the leverage effect measured by the integrated correlation with an additional volatility instrument. Li and Xiu (2016) employ the PCA directly, i.e., imposing the sparsity on eigenvectors (see, e.g., Jolliffe, Trendafilov, and Uddin (2003), Zou, Hastie, and Tibshirani (2006), d’Aspremont, Ghaoui, Jordan, and Lanckriet (2007), Johnstone and Lu (2009), and Amini and Wainwright (2009).) None of these methods are currently applicable to dependent data nor are central limit theorems available for these estimators in a large dimensional setting.
generalized method of moments to investigate structural economic models which include the instantaneous volatility as an explanatory variable. An alternative inference theory in Mykland and Zhang (2009) is designed for a class of estimators based on an aggregation of local estimates using a finite number of blocks.

Also related to the problem we study is the rank inference developed in Jacod, Lejay, and Talay (2008) and Jacod and Podolskij (2013), where the cross-sectional dimension is also fixed and the processes follow continuous Itô semimartingales. Allowing for an increasing dimension but with an additional sparsity structure, Wang and Zou (2010) consider the estimation of integrated covariance matrix with high-dimensional and high frequency data in the presence of measurement errors. Tao, Wang, and Zhou (2013) investigate the minimax rate of convergence for covariance matrix estimation in the same setting; see also Tao, Wang, and Chen (2013) and Tao, Wang, Yao, and Zou (2011) for related work. Zheng and Li (2011) establish a Marcenko-Pastur type theorem for analyzes the spectral distribution of the integrated covariance matrix for a special class of diffusion processes. In a similar setting, Heinrich and Podolskij (2014) study the spectral distribution of empirical covariance matrices of Brownian integrals.

Finally, while the PCA developments in this paper are fully nonparametric without any assumptions on the existence of a factor structure, related techniques to those developed here can be employed to estimate parametric or semiparametric continuous-time models with a factor structure, which is a distinct problem from PCA (just as it is in a low frequency setting); see Aït-Sahalia and Xiu (2015). Independently, Pelger (2015a) and Pelger (2015b) have proposed an alternative estimator for the number of factors and factor loadings.

The paper is organized as follows. Section 2 sets up the notation and the model. Section 3 provides the estimators and the main asymptotic theory. Section 4 reports the results of Monte Carlo simulations. We then implement the method to analyze by PCA the covariance structure of the S&P 100 stocks in Section 5, asking whether at high frequency cross-sectional patterns in stock returns are compatible with the well-established low frequency evidence of a low-dimensional common factor structure (e.g., Fama and French (1993)). We find that, excluding jump variation, three Brownian factors explain between 50 and 60% of continuous variation of the stock returns, that their explanatory power varies over time and that during the recent financial crisis, the first principal component becomes increasingly dominant, explaining up to over 60% of the variation on its own, capturing market-wide, systemic, risk. Despite the differences in methods, time periods, and length of observation, these empirical findings at high frequency are surprisingly consistent with the well-established low frequency Fama-French results of three (or up to five) common factors.

The empirical finding that the low dimensional PC structure that is known to hold at low frequency using decades of daily or lower frequency returns data also holds at high frequency using a week’s worth of data is not obvious, and could not have been obtained without the methodology developed in this paper. More often than not, time aggregation results in patterns that are quite

5
different in data sampled at vastly different frequencies. Of course, the inherent design limitation of PCA is that it does not lend itself naturally to an identification of what the principal components are in terms of economic factors. Nevertheless, we find evidence that the first principal component shares time series characteristics with the overall market return, and, using biplots, we also find that the second principal component drives the common variation of financial sector stocks during the weeks most associated with the recent financial crisis, but not in other weeks. Section 6 concludes. Proofs are in the appendix.

2 The Setup

2.1 Standard Notations and Definitions

In what follows, \( \mathbb{R}_d \) denotes the d-dimensional Euclidean space, and its subset \( \mathbb{R}_d^+ \) contains non-negative real vectors. Let \( e^k \) be the unit vector in \( \mathbb{R}_d^+ \), with the kth entry equal to 1, \( 1 = \sum_{k=1}^{d} e^k \), and \( I \) be the identity matrix. \( M_d \) denotes the Euclidean space of all \( d \times d \) real-valued symmetric matrices. \( M_d^+ \) is a subset of \( M_d \), which includes all positive-semidefinite matrices. We use \( M_d^{++} \) to denote the set of all positive-definite matrices. The Euclidean space \( M_d \) is equipped with an inner product \( \langle A, B \rangle = \text{Tr}(AB) \). We use \( \| \cdot \| \) to denote the Euclidean norm for vectors or matrices, and the superscript “+” to denote the Moore-Penrose inverse of a real-matrix, see e.g., Magnus and Neudecker (1999).

All vectors are column vectors. The transpose of any matrix \( A \) is denoted by \( A^\top \). The ith row of \( A \) is written as \( A_{i,:} \), and the jth column of \( A \) is \( A_{:,j} \). \( A_{ij} \) denotes the \((i,j)\)th entry of \( A \). The operator \( \text{diag}: M_d \to \mathbb{R}_d \) is defined as \( \text{diag}(A) = (A_{11}, A_{22}, \ldots, A_{dd})^\top \). In addition, we define its inverse operator \( \text{Diag} \), which maps a vector \( x \) to a diagonal matrix.

Let \( f \) be a function from \( \mathbb{R}_d^+ \) to \( \mathbb{R} \). The gradient of \( f \) is written as \( \partial f \), and its Hessian matrix is denoted as \( \partial^2 f \). The derivative of a matrix function \( F: M_d^+ \to \mathbb{R} \) is denoted by \( \partial F \in M_d \), with each element written as \( \partial_{ij} F \), for \( 1 \leq i, j \leq d \). Note that the derivative is defined in the usual sense, so that for any \( A \in M_d^+ \), \( \partial_{ij} A = J_{ij} \), where \( J_{ij} \) is a single-entry matrix with the \((i,j)\)th entry equal to 1. The Hessian matrix of \( F \) is written as \( \partial^2 F \), with each entry referred to as \( \partial_{jk,lm}^2 F \), for \( 1 \leq j, k, l, m \leq d \). We use \( \partial^k \) to denote kth order derivatives and \( \delta_{i,j} \) to denote the Kronecker’s delta function giving 1 if \( i = j \) or 0 otherwise. A function \( f \) is Lipchitz if there exists a constant \( K \) such that \( |f(x+h) - f(x)| \leq K \|h\| \). In the proof, \( K \) is a generic constant which may vary from line to line. A function with \( k \)th continuous derivatives is denoted a \( C^k \) function.

Data are sampled discretely every \( \Delta_n \) units of time. All limits are taken as \( \Delta_n \to 0 \). \( \overset{\text{u.c.p.}}{\to} \) denotes uniformly on compacts in probability, and \( \overset{\text{P}}{\to} \) denotes convergence in probability. We use \( \overset{\text{E}}{\to} \) to denote stable convergence in law. We write \( a_n \overset{\text{u.c.p.}}{\to} b_n \) if for some \( c \geq 1 \), \( b_n/c \leq a_n \leq cb_n \) for all \( n \). Finally, \( [\cdot, \cdot] \) denotes the quadratic covariation between Itô semimartingales, and \([\cdot, \cdot]^c\)
the continuous part thereof.

2.2 Eigenvalues and Eigenvectors

We now collect some preliminary results about eigenvalues and eigenvectors. For any vector $x \in \mathbb{R}^d_+$, $\bar{x}$ denotes the vector with the same entries as $x$, ordered in a non-increasing order. We use $\mathbb{R}^+_d$ to denote the subset of $\mathbb{R}^+_d$ containing vectors $x$ satisfying $x_1 \geq x_2 \geq \ldots \geq x_d \geq 0$.

By convention, for any $x \in \mathbb{R}^+_d$, we can write:

$$x_1 = \ldots = x_{g_1} > x_{g_1+1} = \ldots = x_{g_2} > \ldots x_{g_r-1} > x_{g_r-1+1} = \ldots = x_{g_r} \geq 0,$$

where $g_r = d$, and $r$ is the number of distinct element. $\{g_1, g_2, \ldots, g_r\}$ depends on $x$. We then define a corresponding partition of indices as $I_j = \{g_{j-1} + 1, g_{j-1} + 2, \ldots, g_j\}$, for $j = 1, 2, \ldots, r$.

For any $A \in \mathcal{M}_d^+$, $\lambda(A) = (\lambda_1(A), \lambda_2(A), \ldots, \lambda_d(A))^\top$ is the vector of its eigenvalues in a non-increasing order. This notation also permits us to consider $\lambda$ as a mapping from $\mathcal{M}_d^+$ to $\mathbb{R}^+_d$. An important result establishing the continuity of $\lambda$ is, see, e.g., Tao (2012):

**Lemma 1.** $\lambda : \mathcal{M}_d^+ \rightarrow \mathbb{R}_d^+$ is Lipchitz.

Associated with any eigenvalue $\lambda_g$ of $A \in \mathcal{M}_d^+$, we denote its eigenvector as $\gamma_g$, which satisfies $A\gamma_g = \lambda_g \gamma_g$, and $\gamma_g^\top \gamma_g = 1$. The eigenvector, apart from its sign, is uniquely defined when $\lambda_g$ is simple. In such a case, without loss of generality we require the first non-zero element of the eigenvector to be positive. In the presence of repeated eigenvalues, the eigenvector is determined up to an orthogonal transformation. In any case, we can choose eigenvectors such that for any $g \neq h$, $\gamma_g^\top \gamma_h = 0$, see, e.g., Anderson (1958).

When $\lambda_g$ is a simple root, we regard $\gamma_g$ as another vector-valued function of $A$. It turns out in this case, both $\lambda_g(\cdot)$ and $\gamma_g(\cdot)$ are infinitely smooth at $A$, see, e.g., Magnus and Neudecker (1999).

**Lemma 2.** Suppose $\lambda_g$ is a simple root of $A \in \mathcal{M}_d^+$, then $\lambda_g : \mathcal{M}_d^+ \rightarrow \mathbb{R}_d^+$ and $\gamma_g : \mathcal{M}_d^+ \rightarrow \mathbb{R}_d$ are $C^\infty$ at $A$. Moreover, we have

$$\partial_{jk} \lambda_g(A) = \gamma_{gj}(A) \gamma_{gk}(A), \quad \text{and} \quad \partial_{jk} \gamma_g(A) = (\lambda_g \mathbb{I} - A)_{g_j, g_k}$$

where $\gamma_{gk}$ is the $k$th entry of $\gamma_g$. If in addition all the eigenvalues of $A$ are simple, with $(\gamma_1, \gamma_2, \ldots, \gamma_d)$ being the corresponding eigenvectors, then

$$\partial_{jk, lm}^2 \gamma_{gh} = \sum_{p \neq g} \frac{1}{(\lambda_g - \lambda_p)^2} \left( \gamma_{pl} \gamma_{gm} \gamma_{ph} \gamma_{pj} \gamma_{gk} - \gamma_{pl} \gamma_{pm} \gamma_{ph} \gamma_{pj} \gamma_{gk} \right)$$

$$+ \sum_{p \neq g} \sum_{q \neq p} \frac{1}{(\lambda_g - \lambda_p)(\lambda_g - \lambda_q)} \gamma_{ql} \gamma_{pm} \gamma_{qh} \gamma_{pj} \gamma_{gk}$$
In general, while the eigenvalue is always a continuous function, the eigenvector associated with a repeated root is not necessarily continuous. In what follows, we only consider estimation of an eigenvector when it is associated with a simple eigenvalue.

### 2.3 Dynamics of the Variable

The process we analyze is a general \( d \)-dimensional Itô semimartingale, defined on a filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\) with the following Grigelionis representation:

\[
X_t = X_0 + \int_0^t b_s \, ds + \int_0^t \sigma_s \, dW_s + (\delta 1_{\{\delta \leq 1\}}) * (\mu - \nu)_t + (\delta 1_{\{\delta > 1\}}) * \mu_t,
\]

where \( W \) is a \( d \)-dimensional Brownian motion, \( \mu \) is a Poisson random measure on \( \mathbb{R}^+ \times \mathbb{R}_d \) with the compensator \( \nu(dt, dx) = dt \otimes \tilde{\nu}(dx) \), and \( \tilde{\nu} \) is a \( \sigma \)-finite measure. More details on high frequency models and asymptotics can be found in the book Aït-Sahalia and Jacod (2014).

The volatility process \( \sigma_s \) is càdlàg, \( c_s = (\sigma \sigma^\top)_s \in \mathcal{M}_d^+ \), for any \( 0 \leq s \leq t \). We denote \( d \) non-negative eigenvalues of \( c_s \) by \( \lambda_{1,s} \geq \lambda_{2,s} \geq \ldots \geq \lambda_{d,s} \), summarized in a vector \( \lambda_s \). As is discussed above, we sometimes write \( \lambda_s = \lambda(c_s) \), regarding \( \lambda(\cdot) \) as a function of \( c_s \). By Lemma 1, \( \lambda(\cdot) \) is a continuous function so that \( \lambda(c_s) \) is càdlàg.

### 2.4 Principal Component Analysis at High Frequency

Similar to the classical PCA, the PCA procedure in this setting consists in searching repeatedly for instantaneous linear combinations of \( X \), namely principal components, which maximize certain measure of variation, while being orthogonal to the principal components already constructed, at any time between 0 and \( t \). In contrast to the classical setting, where one maximizes the variance (see, e.g., Anderson (1958)), the criterion here is the continuous part of the quadratic variation.

**Lemma 3.** Suppose that \( X \) is a \( d \)-dimensional vector-valued process described in (2). Then there exists a sequence of \( \{\lambda_{g,s}, \gamma_{g,s}\}_{1 \leq g \leq d, 0 \leq s \leq t} \) such that

\[
c_s \gamma_{g,s} = \lambda_{g,s} \gamma_{g,s}, \quad \gamma_{g,s}^\top \gamma_{g,s} = 1, \quad \text{and} \quad \gamma_{h,s}^\top c_s \gamma_{g,s} = 0,
\]

where \( \lambda_{1,s} \geq \lambda_{2,s} \geq \ldots \geq \lambda_{d,s} \geq 0 \). Moreover, for any càdlàg and vector-valued adapted process \( \gamma_s \),
such that $\gamma_s^T \gamma_s = 1$, and $\gamma_h^T c_g \gamma_{h,s} = 0$, $1 \leq h \leq g - 1$,  
\[
\int_0^u \lambda_{g,s} ds \geq \left[ \int_0^u \gamma_{s-}^T dX_s, \int_0^u \gamma_{s-}^T dX_s \right]^c, \quad \text{for any } 0 \leq u \leq t.
\]

When $\lambda_{g,s}$ is a simple root of $c_g$ between $0$ and $t$, then $\gamma_{g,s}$ is an adapted càdlàg process, due to the continuity of $\gamma_g(\cdot)$ by Lemma 2, so that we can construct its corresponding principal component $\int_0^t \gamma_{g,s}^T dX_s$.

Given Lemma 3, we can define our parameters of interest: the realized eigenvalue, i.e. $\int_0^t \lambda_{g,s} ds$; the realized principal components associated with some simple root $\lambda_g$, i.e. $\int_0^t \gamma_{g,s}^T dX_s$. It may also be worth investigating the average loading on the principal component, i.e. the realized eigenvector, which is $\int_0^t \gamma_{g,s} ds$. Since the definitions of these quantities all rely on integrals, we call them integrated eigenvalues, integrated principal components, and integrated eigenvectors.

The above analysis naturally leads us to consider inference for $\int_0^t \lambda(c_g) ds$, for which the differentiability of $\lambda(\cdot)$ is critical: we start with the convergence of an estimator $\hat{c}_g$ to $c_g$, from which we need in a delta method sense to infer the convergence of the integrated eigenvalues estimator. A simple eigenvalue is $C^\infty$-differentiable, but for repeated eigenvalues this is not necessarily the case. For this reason, in order to fully address the estimation problem, we need to introduce spectral functions.

**2.5 Spectral Functions**

A real-valued function $F$ defined on a subset of $\mathcal{M}_d^+$ is called a spectral function (see, e.g., Friedland (1981)) if for any orthogonal matrix $O$ in $\mathcal{M}_d$ and $X$ in $\mathcal{M}_d^+$, $F(X) = F(O^T X O)$. We describe sets in $\mathbb{R}_d^+$ and functions from $\mathbb{R}_d^+$ to $\mathbb{R}^+$ as symmetric if they are invariant under coordinate permutations. That is, for any symmetric function $f$ with an open symmetric domain in $\mathbb{R}_d^+$, we have $f(x) = f(Px)$ for any permutation matrix $P \in \mathcal{M}_d$. Associated with any spectral function $F$, we define a function $f$ on $\mathbb{R}_d^+$, so that $f(x) = F(\text{Diag}(x))$. Since permutation matrices are orthogonal, $f$ is symmetric, and $f \circ \lambda = F$, where $\circ$ denotes function composition.

We will need to differentiate the spectral function $F$. We introduce a matrix function associated with the corresponding symmetric function $f$ of $F$, which, for any $x \in \mathbb{R}_d^+$ in the form of (1), is given by

\[
\mathcal{A}_f^{p,q}(x) = \begin{cases} 
0 & \text{if } p = q; \\
\partial^2_{pp} f(x) - \partial^2_{qq} f(x) & \text{if } p \neq q, \quad \text{and } p, q \in I_l; \\
(\partial_p f(x) - \partial_q f(x))/(x_p - x_q) & \text{otherwise.}
\end{cases}
\]

for some $l = \{1, 2, \ldots, r\}$.

The following lemma collects some known and useful results regarding the continuous differentiability and convexity of spectral functions, which we will use below:
Lemma 4. The symmetric function \( f \) is twice continuously differentiable at a point \( \lambda(A) \in \mathbb{R}_d^+ \) if and only if the spectral function \( F = f \circ \lambda \) is twice continuously differentiable at the point \( A \in \mathcal{M}_d^+ \). The gradient and the Hessian matrix are given below:

\[
\partial_{jk}(f \circ \lambda)(A) = \sum_{p=1}^{d} O_{pj} \partial_p f(\lambda(A)) O_{pk},
\]

\[
\partial_{jk,lm}^2(f \circ \lambda)(A) = \sum_{p,q=1}^{d} \partial_{pq}^2 f(\lambda(A)) O_{pl} O_{pm} O_{qj} O_{qk} + \sum_{p,q=1}^{d} A_{pq}^2(\lambda(A)) O_{pl} O_{pj} O_{qk} O_{qm},
\]

where \( O \) is any orthogonal matrix that satisfies \( A = O^T \text{Diag}(\lambda(A)) O \). More generally, \( f \) is a \( C^k \) function at \( \lambda(A) \) if and only if \( F \) is \( C^k \) at \( A \), for any \( k = 0, 1, \ldots, \infty \). In addition, \( f \) is a convex function if and only if \( F \) is convex.

Next, we note that both simple and repeated eigenvalues can be viewed as special cases of spectral functions.

Example 1 (A Simple Eigenvalue): Suppose the \( k \)th eigenvalue of \( A \in \mathcal{M}_d^+ \) is simple, that is, \( \lambda_{k-1}(A) > \lambda_k(A) > \lambda_{k+1}(A) \). Define for any \( x \in \mathbb{R}_d^+ \),

\[
f(x) = \text{the } k \text{th largest entry in } x = \bar{x}_k.
\]

Apparently, \( f \) is a symmetric function, and it is \( C^\infty \) at any point \( y \in \mathbb{R}_d^+ \), with \( y_{k-1} > y_k > y_{k+1} \). Indeed, \( \partial f(y) = e^k \), and \( \partial^l f(y) = 0 \) for \( l \geq 2 \). By Lemma 4, \( \lambda_k(A) = (f \circ \lambda)(A) \) is \( C^\infty \) at \( A \).

Example 2 (Non-Simple Eigenvalues): Suppose the eigenvalues of \( A \in \mathcal{M}_d^+ \) satisfy:

\[
\lambda_{g_1-1}(A) > \lambda_{g_1-1+1}(A) \geq \ldots \geq \lambda_{g_l}(A) > \lambda_{g_l+1}(A),
\]

for some \( 1 \leq g_1 < g_l \leq d \). By convention, when \( g_l = d \), the last “\( > \)” above is not used. Consider the following function \( f \), evaluated at \( x \in \mathbb{R}_d^+ \), which is given by

\[
f(x) = \frac{1}{g_l - g_{l-1}} \sum_{j=g_{l-1}+1}^{g_l} \bar{x}_j.
\]

It is easy to verify that \( f \) is symmetric, and \( C^\infty \) at any point \( y \in \mathbb{R}_d^+ \) that satisfies \( y_{g_{l-1}} > y_{g_{l-1}+1} \geq \ldots \geq y_{g_l} > y_{g_l+1} \). Moreover, \( \partial f(y) = \frac{1}{g_l - g_{l-1}} \sum_{k=g_{l-1}+1}^{g_l} e^k \), and \( \partial^l f(y) = 0 \), for any \( l \geq 2 \). As a result of Lemma 4, the corresponding spectral function,

\[
F(A) = (f \circ \lambda)(A) = \frac{1}{g_l - g_{l-1}} \sum_{j=g_{l-1}+1}^{g_l} \lambda_j(A),
\]
is $C^\infty$ at $A$. In the special case where

$$\lambda_{g_{l-1}}(A) > \lambda_{g_{l-1}+1}(A) = \ldots = \lambda_{g_l}(A) > \lambda_{g_{l+1}}(A),$$

i.e., there is a repeated eigenvalue, and $F(A) = \lambda_{g_{l-1}+1}(A) = \ldots = \lambda_{g_l}(A)$. By contrast, $\lambda_j(A)$ is not differentiable, for any $g_{l-1} + 1 \leq j \leq g_l$.

**Example 3 (Trace and Determinant):** For any $x \in \mathbb{R}_d^+$, define

$$f_1(x) = \sum_{j=1}^d x_j, \quad \text{and} \quad f_2(x) = \prod_{j=1}^d x_j.$$

Both $f_1$ and $f_2$ are symmetric and $C^\infty$. Therefore, for any $A \in \mathcal{M}_d^+$, $\text{Tr}(A) = (f_1 \circ \lambda)(A)$ and $\det(A) = (f_2 \circ \lambda)(A)$ are $C^\infty$ at $A$.

The previous examples make it clear that the objects of interest, eigenvalues, are special cases of spectral functions, whether they are simple or repeated. (We are also able to estimate the trace and determinant “for free”.) Lemma 4 links the differentiability of a spectral function $F$, which is needed for statistical inference, to that of its associated symmetric function $f$. The key advantage is that differentiability of $f$ is easier to establish.

Before turning to inference, we prove a useful result that characterizes the topology of the set of matrices with special eigenvalue structures. The domain of spectral functions we consider will be confined to this set in which these spectral functions are smooth.

**Lemma 5.** For any $1 \leq g_1 < g_2 < \ldots < g_r \leq d$, the set

$$\mathcal{M}(g_1, g_2, \ldots, g_r) = \{ A \in \mathcal{M}_d^+ \mid \lambda_{g_l}(A) > \lambda_{g_{l+1}}(A), \text{ for any } l = 1, 2, \ldots, r - 1 \}$$  \hspace{1cm} (3)

is dense and open in $\mathcal{M}_d^+$. In particular, the set of positive-definite matrices with distinct eigenvalues, i.e., $\mathcal{M}(1, 2, \ldots, d)$, is dense and open in $\mathcal{M}_d^+$.  

We conclude this section with some additional notations. First, we introduce a symmetric open subset of $\mathbb{R}_d^+/\{0\}$, the image of $\mathcal{M}(g_1, g_2, \ldots, g_r)$ under $\lambda(\cdot)$:

$$D(g_1, g_2, \ldots, g_r) = \{ x \in \mathbb{R}_d^+/\{0\} \mid \bar{x}_{g_l} > \bar{x}_{g_{l+1}}, \text{ for any } l = 1, 2, \ldots, r - 1 \}. \hspace{1cm} (4)$$

We also introduce a subset of $\mathcal{M}(g_1, g_2, \ldots, g_r)$, in which our spot covariance matrix $c_t$ will take values.

$$\mathcal{M}^*(g_1, g_2, \ldots, g_r) = \{ A \in \mathcal{M}_d^+ \mid$$

11
\[ \lambda_1(A) = \ldots = \lambda_{g_1}(A) > \lambda_{g_1+1}(A) = \ldots = \lambda_{g_2}(A) > \ldots \lambda_{g_{r-1}}(A) > \lambda_{g_{r-1}+1}(A) = \ldots = \lambda_{g_r}(A) \] 

Finally, we introduce the following set, which becomes relevant if we are only interested in a simple eigenvalue \( \lambda_g \):
\[ \mathcal{M}(g) = \{ A \in \mathcal{M}_d^{++} | \lambda_{g-1}(A) > \lambda_g(A) > \lambda_{g+1}(A) \} \text{, and } \mathcal{D}(g) = \lambda(\mathcal{M}(g)). \]

By convention, we ignore the first (resp. second) inequality in the definition of \( \mathcal{M}(g) \) when \( g = 1 \) (resp. \( g = d \)).

3 Estimators and Asymptotic Theory

3.1 Assumptions

We start with the standard assumption on the process \( X \):²

**Assumption 1.** The drift term \( b_t \) is progressively measurable and locally bounded. The spot covariance matrix \( c_t = (\sigma^\top)_t \) is an Itô semimartingale. Moreover, for some \( \gamma \in [0,1) \), there is a sequence of stopping times \( (\tau_n) \) increasing to \( \infty \), and a deterministic function \( \delta_n \) such that
\[ \int_{\mathbb{R}_d} \delta_n(x)^\gamma \nu(dx) < \infty \text{ and that } \|\delta(\omega,t,x)\| \wedge 1 \leq \delta_n(x), \text{ for all } (\omega,t,x) \text{ with } t \leq \tau_n(\omega). \]

In view of the previous examples, our main theory is tailored to the statistical inference on \( \int_0^t F(c_s)ds \). Thanks to Lemma 4, we can make assumptions directly on \( f \) instead of \( F \), which are much easier to verify.

**Assumption 2.** Suppose \( F \) is a vector-valued spectral function, and \( f \) is the corresponding vector-valued symmetric function such that \( F = f \circ \lambda \). \( f \) is a continuous function, and satisfies \( \|f(x)\| \leq K(1 + \|x\|^{\zeta}) \), for some \( \zeta > 0 \).

In all the examples of Section 2.5, this assumption holds. By Lemma 1 and Assumption 2, \( F(c_s) \) is càdlàg, and the integral \( \int_0^t F(c_s)ds \) is well-defined.

The above assumptions are sufficient to ensure the desired consistency of estimators we will propose. Additional assumptions are required to establish their asymptotic distribution.

**Assumption 3.** There exists some open and convex set \( \mathcal{C} \), such that its closure \( \bar{\mathcal{C}} \subset \mathcal{M}(g_1,g_2,\ldots,g_r) \), where \( 1 \leq g_1 < g_2 < \ldots < g_r \leq d \), and that for any \( 0 \leq s \leq t \), \( c_s \in \mathcal{C} \cap \mathcal{M}^*(g_1,g_2,\ldots,g_r) \). Moreover, \( f \) is \( C^3 \) on \( \mathcal{D}(g_1,g_2,\ldots,g_r) \).

Assumption 3, in particular, \( c_s \in \mathcal{M}^*(g_1,g_2,\ldots,g_r) \), guarantees that different groups of eigenvalues do not cross over within \([0,t]\). This condition is mainly used to deliver the joint central limit

²Using the notation on Page 583 of Jacod and Protter (2012), Assumption 1 states that the process \( X \) satisfies Assumption (H-1) and that \( c_t \) satisfies Assumption (H-2).
Theorem for spectral functions that depend on all eigenvalues, although it may not be necessary for some special cases, such as $\det(\cdot)$ and $\text{Tr}(\cdot)$, which are smooth everywhere. The convexity condition on $C$ is in principle not difficult to satisfy, given that $\mathcal{M}(g_1, g_2, \ldots, g_r)$ can be embedded into some Euclidean space of real vectors. This condition is imposed to ensure that the domain of the spectral function can be restricted to certain neighborhood of $\{c_s\}_{0 \leq s \leq t}$, in which the function is smooth and the mean-value theorem can be applied. This assumption is not needed if $c_t$ is continuous.

It is also worth mentioning that all eigenvalues of $c_t$ being distinct is a special case, which is perhaps the most relevant scenario in practice, as is clear from Lemma 5. Even in this scenario, Assumption 3 is required (with $g_1, g_2, \ldots, g_r$ being chosen as $1, 2, \ldots, d$), because the eigenvalue function $\lambda(\cdot)$ is not everywhere differentiable.

Assumption 3 resembles the spacial localization assumption in Li, Todorov, and Tauchen (2014) and Li and Xiu (2016), which is different from the polynomial growth conditions proposed by Jacod and Rosenbaum (2013). The growth conditions are not satisfied in our setting, when the difference between two groups of repeated eigenvalues approaches zero.

If we are only interested in the spectral function that depends on one simple eigenvalue, e.g., the largest and simple integrated eigenvalue, $\int_0^t \lambda_1(c_s) ds$, then it is only necessary to ensure that $\lambda_1(c_s) > \lambda_2(c_s)$, for $0 \leq s \leq t$, regardless of whether the remaining eigenvalues are simple or not. We thereby introduce the following assumption for this scenario, which is much weaker than Assumption 3.

**Assumption 4.** There exists some open and convex set $C$, such that $\bar{C} \subset \mathcal{M}(g)$, for some $g = 1, 2, \ldots, d$, and that for any $0 \leq s \leq t$, $c_s \in C$. Moreover, $f$ is $C^3$ on $\mathcal{D}(g)$.

### 3.2 Realized Spectral Functions

We now turn to the construction of the estimators. To estimate the integrated spectral function, we start with estimation of the spot covariance matrix. Suppose we have equidistant observations on $X$ over the interval $[0, t]$, separated by a time interval $\Delta_n$. We form non-overlapping blocks of length $k_n \Delta_n$. At each $ik_n \Delta_n$, we estimate $c_{ik_n} \Delta_n$ by

$$\hat{c}_{ik_n} \Delta_n = \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} (\Delta_{ik_n+j} X_i^T \Delta_{ik_n+j} X_i) \mathbb{1}_\{\|\Delta_{ik_n+j} X_i\| \leq u_n\}^\top,$$

where $u_n = \alpha \Delta_n^{\overline{\alpha}}$, and $\Delta_n X_i = X_i(\Delta_n) - X_i(\Delta_n)$. Choices of $\alpha$ and $\overline{\alpha}$ are standard in the literature (see, e.g., Aït-Sahalia and Jacod (2014)) and are discussed below when implemented in simulations.

We then estimate eigenvalues of $\hat{c}_{ik_n} \Delta_n$ by solving for the roots of $|\hat{c}_{ik_n} \Delta_n - \lambda| = 0$. Using the notation in Lemma 1, we have $\lambda(\hat{c}_{ik_n} \Delta_n) = \hat{\lambda}_{ik_n} \Delta_n$. Almost surely, the eigenvalues stacked in $\hat{\lambda}_{ik_n} \Delta_n$ are distinct (see, e.g., Okamoto (1973)) so that we have $\hat{\lambda}_1_{ik_n} \Delta_n > \hat{\lambda}_2_{ik_n} \Delta_n > \ldots > \hat{\lambda}_d_{ik_n} \Delta_n$. Our
proposed estimator of the integrated spectral function is then given by\(^3\)

\[
V(\Delta_n, X; F) = k_n \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} f(\hat{\lambda}_{ik_n \Delta_n}).
\]  

(6)

We start by establishing consistency of this estimator:

**Theorem 1.** Suppose Assumptions 1 and 2 hold. Also, either \(\zeta \leq 1\), or \(\zeta > 1\) with \(\varpi \in \left[\frac{(\zeta-1)}{2}, \frac{1}{2}\right]\) holds. Then the estimator (6) is consistent. As \(k_n \to \infty\) and \(k_n \Delta_n \to 0\),

\[
V(\Delta_n, X; F) \overset{u.c.p.}{\to} \int_0^t F(c_s)ds.
\]  

(7)

Next, obtaining a central limit theorem for the estimator is more involved, as there is a second-order asymptotic bias associated with the estimator (6), a complication that is also encountered in other situations such as the quarticity estimator in Jacod and Rosenbaum (2013), the GMM estimator in Li and Xiu (2016), or the leverage effect estimator by Kalnina and Xiu (2013). The bias here is characterized as follows:

**Proposition 1.** Suppose Assumptions 1, 2, and 3 hold. In addition, \(k_n \asymp \Delta_n^{-\zeta}\) and \(u_n \asymp \Delta_n^{\varpi}\) for some \(\zeta \in \left(\frac{1}{2}, \frac{1}{2}\right)\) and \(\varpi \in \left[\frac{1-\zeta}{2}, \frac{1}{2}\right]\). As \(\Delta_n \to 0\), we have

\[
k_n \left( V(\Delta_n, X; F) - \int_0^t F(c_s)ds \right) \overset{p}{\to} \frac{1}{2} \sum_{j,k,l,m=1}^d \int_0^t \partial_{j,k,lm}^2 F(c_s) (c_{jl,s} c_{km,s} + c_{jm,s} c_{kl,s}) ds.
\]  

(8)

The characterization of the bias in (8) suggests a bias-corrected estimator as follows:

\[
\tilde{V}(\Delta_n, X; F) = k_n \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} \left\{ F(\hat{\lambda}_{ik_n \Delta_n}) \right\}
\]

\-

\[
- \frac{1}{2k_n} \sum_{j,k,l,m=1}^d \partial_{j,k,lm}^2 F(\hat{\lambda}_{ik_n \Delta_n}) (\hat{\lambda}_{jl,ik_n \Delta_n} \hat{\lambda}_{km,ik_n \Delta_n} + \hat{\lambda}_{jm,ik_n \Delta_n} \hat{\lambda}_{kl,ik_n \Delta_n})
\]  

(9)

We then derive the asymptotic distribution of the bias-corrected estimator:

**Theorem 2.** Suppose Assumptions 1, 2, and 3 hold. In addition, \(k_n \asymp \Delta_n^{-\zeta}\) and \(u_n \asymp \Delta_n^{\varpi}\) for some \(\zeta \in \left(\frac{1}{2}, \frac{1}{3}\right)\) and \(\varpi \in \left[\frac{1-\zeta}{2}, \frac{1}{2}\right]\). As \(\Delta_n \to 0\), we have

\[
\frac{1}{\sqrt{\Delta_n}} \left( \tilde{V}(\Delta_n, X; F) - \int_0^t F(c_s)ds \right) \overset{d}{\to} W_t.
\]  

(10)

\(^3\)We prefer this estimator to the alternative one using overlapping windows, because the overlapping implementation runs much slower. In a finite sample, both estimators have a decent performance.
where \( W \) is a continuous process defined on an extension of the original probability space, which conditionally on \( \mathcal{F} \), is continuous centered Gaussian martingale with its covariance given by

\[
E(W_{p,t}W_{q,t} | \mathcal{F}) = \int_0^t \sum_{j,k,l,m=1}^d \partial_{jk} F_p(c_s) \partial_{lm} F_q(c_s) \left( c_{jl,s}c_{km,s} + c_{jm,s}c_{kl,s} \right) ds.
\]  

(11)

Remark 1. As previously noted, in the case when the spectral function \( F \) only depends on a simple eigenvalue \( \lambda_g \), the same result holds under the weaker Assumption 4 instead of 3.

A feasible implementation of this distribution requires an estimator of the asymptotic variance, which we construct as follows:

Proposition 2. The asymptotic variance of \( \tilde{V}(\Delta_n, X; F) \) can be estimated consistently by:

\[
k_n \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} \sum_{j,k,l,m=1}^d \partial_{jk} F_p(c_s) \partial_{lm} F_q(c_s) \left( c_{jl,s}c_{km,s} + c_{jm,s}c_{kl,s} \right) ds.
\]

(12)

3.3 Realized Eigenvalues

We next specialize the previous theorem to obtain a central limit theorem for the realized eigenvalue estimators. With the structure of eigenvalues in mind, we use a particular vector-valued spectral function \( F^\lambda \), tailor-made to deliver the asymptotic theory we need for realized eigenvalues:

\[
F^\lambda(\cdot) = \left( \frac{1}{g_1} \sum_{j=1}^{g_1} \lambda_j(\cdot), \frac{1}{g_2 - g_1} \sum_{j=g_1 + 1}^{g_2} \lambda_j(\cdot), \ldots, \frac{1}{g_r - g_{r-1}} \sum_{j=g_{r-1} + 1}^{g_r} \lambda_j(\cdot) \right)^T.
\]

(13)

Apparently, if a group contains only one \( \lambda_j \), then the corresponding entry of \( F^\lambda \) is equal to this single eigenvalue; if within certain group, all eigenvalues are identical, then the corresponding entry of \( F^\lambda \) yields the common eigenvalue of the group.

Corollary 1. Suppose \( k_n \sim \Delta_n^{-\varsigma} \) and \( u_n \sim \Delta_n^{-\varpi} \) for some \( \varsigma \in \left( \frac{1}{2}, \frac{1}{3} \right) \) and \( \varpi \in \left( \frac{1-\varsigma}{2-\varsigma}, \frac{1}{2} \right) \).

(i) Under Assumption 1, the estimator of integrated eigenvalue vector given by

\[
V(\Delta_n, X; \lambda) = k_n \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} \lambda(\tilde{c}_{ik_n \Delta_n})
\]

is consistent.

(ii) If Assumption 3 further holds, the estimator corresponding to the \( p \)th entry of \( F^\lambda \) given by (9)
can be written explicitly as:

\[
\tilde{V}(\Delta_n, X; F_\text{p}) = \frac{k_n \Delta_p}{g_p - g_{p-1}} \sum_{i=0}^{[t/(k_n \Delta_n)]} \sum_{h=0}^{g_p} \left\{ \hat{\lambda}_{h,i} \Delta_n - \frac{1}{k_n} \text{Tr} \left( (\hat{\lambda}_{h,i} \Delta_n - \overline{c}_{ikn} \Delta_n)^+ \overline{c}_{ikn} \Delta_n \right) \hat{\lambda}_{h,i} \Delta_n \right\}.
\]

The joint central limit theorem for \( \tilde{V}(\Delta_n, X; F_\text{p}) = (\tilde{V}(\Delta_n, X; F_1), \ldots, \tilde{V}(\Delta_n, X; F_r))^\top \) is given by

\[
\frac{1}{\sqrt{\Delta_n}} \left( \tilde{V}(\Delta_n, X; F_\text{p}) - \int_0^t F^\lambda(c_s) ds \right) \overset{\mathcal{L}}{\longrightarrow} \mathcal{W}_t^\lambda,
\]

where \( \mathcal{W}_t^\lambda \) is a continuous process defined on an extension of the original probability space, which conditionally on \( \mathcal{F} \), is continuous centered Gaussian martingale with a diagonal covariance matrix given by

\[
\mathbb{E}(\mathcal{W}_t^\lambda(\mathcal{W}_t^\lambda)^\top | \mathcal{F}) = \begin{pmatrix}
\frac{2}{g_1} \int_0^t \lambda_{g_1}^2 ds \\
\frac{2}{g_2 - g_1} \int_0^t \lambda_{g_2}^2 ds \\
\ddots \\
\frac{2}{g_{r-1} - g_{r-2}} \int_0^t \lambda_{g_{r-1}}^2 ds
\end{pmatrix},
\]

where \( F^\lambda(c_s) = (\lambda_{g_1}, \lambda_{g_2}, \ldots, \lambda_{g_{r-1}})^\top \).

(iii) Under Assumptions 1 and 4, our estimator with respect to the \( g \)th simple eigenvalue \( \lambda_g(\cdot) \), is given by

\[
\tilde{V}(\Delta_n, X; \lambda_g) = k_n \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} \left\{ \hat{\lambda}_{g,i} \Delta_n - \frac{1}{k_n} \text{Tr} \left( (\hat{\lambda}_{g,i} \Delta_n - \overline{c}_{ikn} \Delta_n)^+ \overline{c}_{ikn} \Delta_n \right) \hat{\lambda}_{g,i} \Delta_n \right\},
\]

which satisfies:

\[
\frac{1}{\sqrt{\Delta_n}} \left( \tilde{V}(\Delta_n, X; \lambda_g) - \int_0^t F^\lambda(c_s) ds \right) \overset{\mathcal{L}}{\longrightarrow} \mathcal{W}_t^\lambda_g,
\]

where \( \mathcal{W}_t^\lambda_g \) is a continuous process defined on an extension of the original probability space, which conditionally on \( \mathcal{F} \), is a continuous centered Gaussian martingale with its variance \( \int_0^t \lambda_{g,s}^2 ds \).

Remark 2. Corollary 1 is the analogue in our context to the classical results on asymptotic theory for PCA by Anderson (1963), who showed that

\[
\sqrt{n}(\hat{\lambda} - \lambda) \overset{d}{\rightarrow} N(0, 2 \text{Diag}(\lambda_1^2, \lambda_2^2, \ldots, \lambda_d^2)).
\]

where \( \hat{\lambda} \) and \( \lambda \) are the vectors of eigenvalues of the sample and population covariance matrices and \( \lambda \) is simple. Note the similarity between (19) and (16) in the special case where the eigenvalues are simple (in which case \( g_i = i \) and are constant instead of being stochastic. The classical
setting requires the strong assumption that the covariance matrix follows a Wishart Distribution. By contrast, our results are fully nonparametric, relying instead on high frequency asymptotics.

**Remark 3.** If there are eigenvalues of \( c_s \) that are equal to 0 for any \( 0 \leq s \leq t \), then our estimator is super-efficient. This is an interesting case as the rank of the covariance matrix is determined by the number of non-zero eigenvalues, see, e.g., the rank inference in Jacod, Lejay, and Talay (2008) and Jacod and Podolskij (2013).

### 3.4 Realized Principal Components

As we have remarked before, when eigenvalues are simple, the corresponding eigenvectors are uniquely determined. Therefore, we can estimate the instantaneous eigenvectors together with eigenvalues, which eventually leads us to construct the corresponding principal component:

**Proposition 3.** Suppose Assumptions 1 and 4 hold. In addition, \( \gamma_{g,s} \) is a vector-valued function that corresponds to the eigenvector of \( c_s \) with respect to a simple root \( \lambda_{g,s} \). Suppose \( k_n \sim \Delta_n^{-\alpha} \) and \( u_n \sim \Delta_n^\omega \) for some \( \alpha \in \left( \frac{1}{2}, \frac{1}{2} \right) \) and \( \omega \in \left[ \frac{1}{2} - \frac{\alpha}{2}, \frac{1}{2} \right] \). Then we have as \( \Delta_n \to 0 \)

\[
\sum_{i=1}^{[t/(k_n\Delta_n)]-1} \gamma_{g,(i-1)k_n\Delta_n} (X_{(i+1)k_n\Delta_n} - X_{ik_n\Delta_n}) \overset{u.c.p}{\longrightarrow} \int_0^t \gamma_{g,s}^\top dX_s.
\]

### 3.5 Realized Eigenvectors

So far we have constructed the principal components and estimated integrated eigenvalues. It remains to figure out the loading of each entry of \( X \) on each principal component, i.e., the eigenvector. As the eigenvector is stochastic, we estimate the integrated eigenvector. Apart from its sign, an eigenvector is uniquely identified if its associated eigenvalue is simple. We determine the sign hence identify the eigenvector, by requiring a priori certain entry of the eigenvector to be positive, e.g., the first non-zero entry.

**Corollary 2.** Suppose Assumptions 1 and 4 hold. In addition, \( \gamma_{g,s} \) is a vector-valued function that corresponds to the eigenvector of \( c_s \) with respect to a simple root \( \lambda_{g,s} \), for each \( s \in [0,t] \). Then we have as \( \Delta_n \to 0 \)

\[
\frac{1}{\sqrt{\Delta_n}} \left( k_n \Delta_n \sum_{i=0}^{[t/(k_n\Delta_n)]} \left( \gamma_{g,ik_n\Delta_n} + \frac{1}{2k_n} \sum_{p \neq g} \frac{\lambda_{g,ik_n\Delta_n} \lambda_{p,ik_n\Delta_n}}{(\lambda_{g,ik_n\Delta_n} - \lambda_{p,ik_n\Delta_n})^2} \gamma_{g,ik_n\Delta_n} \right) - \int_0^t \gamma_{g,s}^\top ds \right) \overset{u.c.}{\longrightarrow} W_t^\gamma,
\]

where \( W_t^\gamma \) is a continuous process defined on an extension of the original probability space, which conditionally on \( \mathcal{F} \), is a continuous centered Gaussian martingale with its covariance matrix given by

\[
E(W_t^\gamma (W_t^\gamma)^\top | \mathcal{F}) = \int_0^t \lambda_{g,s} (\lambda_{g,s} - c_s)^+ c_s (\lambda_{g,s} - c_s)^+ ds.
\]

17
3.6 “PCA” on the Integrated Covariance Matrix

One may wonder why we chose to estimate the integrated eigenvalues of the spot covariance matrix rather than the eigenvalues of the integrated covariance matrix. Because eigenvalues are complicated functionals of the covariance matrix, the two quantities are not much related, and it turns out that the latter procedure is less informative than the former. For instance, the rank of the instantaneous covariance matrix is determined by the number of non-zero instantaneous eigenvalues. The eigenstructure of the integrated covariance matrix is rather opaque due to the aggregation of instantaneous covariances. Nevertheless, it is possible to construct a “PCA” procedure on the integrated covariance matrix, with the following results:

**Proposition 4.** Suppose the eigenvalues of the integrated covariance matrix $\int_0^t c_s ds$ are given by

$$\lambda_1 = \ldots = \lambda_g > \lambda_{g+1} = \ldots = \lambda_{g+2} > \ldots \lambda_{g+r-1} > \lambda_{g+r-1+1} = \ldots = \lambda_g > 0, \quad \text{for} \ 1 \leq r \leq d.$$

Then, we have, for $\omega \in [1/(4 - 2\gamma), 1/2)$,

$$\frac{1}{\sqrt{n}} \left( F^\lambda \left( \sum_{i=0}^{[t/\Delta_n]} (\Delta_i^n X)(\Delta_i^n X)^\top \{\|\Delta_i^n X\| \leq \alpha \Delta_n^2 \} \right) - F^\lambda \left( \int_0^t c_s ds \right) \right) \overset{\mathcal{L}}{\rightarrow} \tilde{W}^\lambda_t,$$

where $\tilde{W}^\lambda_t$ is a continuous process defined on an extension of the original probability space, which conditionally on $\mathcal{F}$, is a continuous centered Gaussian martingale with its covariance matrix given by

$$\mathbb{E}(\tilde{W}^\lambda_{i,t}\tilde{W}^\lambda_{j,t}|\mathcal{F}) = \sum_{u,v,k,l} \partial_{uv} F^\lambda_i \left( \int_0^t c_s ds \right) \left( \int_0^t (c_{uk,s}c_{vl,s} + c_{ul,s}c_{vk,s}) ds \right) \partial_{kl} F^\lambda_j \left( \int_0^t c_s ds \right),$$

$$\partial_{uv} F^\lambda_i (A) = \frac{1}{g_i - g_{i-1}} \sum_{k=g_{i-1}+1}^{g_i} O_{ku} O_{kv},$$

where $O$ is any orthogonal matrix that satisfies $A = O^\top \text{Diag}(\lambda(A)) O$.

Similarly, we have:

**Proposition 5.** Suppose $\omega \in [1/(4 - 2\gamma), 1/2)$. For the eigenvector $\gamma_g$ corresponding to some simple eigenvalue $\lambda_g$ of $\int_0^t c_s ds$,

$$\frac{1}{\sqrt{n}} \left( \gamma_g \left( \sum_{i=0}^{[t/\Delta_n]} (\Delta_i^n X)(\Delta_i^n X)^\top \{\|\Delta_i^n X\| \leq \alpha \Delta_n^2 \} \right) - \gamma_g \left( \int_0^t c_s ds \right) \right) \overset{\mathcal{L}}{\rightarrow} \tilde{W}^\gamma_t,$$

where $\tilde{W}^\gamma_t$ is a continuous process defined on an extension of the original probability space, which conditionally on $\mathcal{F}$, is a continuous centered Gaussian martingale with its covariance matrix given...
by

\[ E(\hat{W}_{i,t}^{\gamma}(\hat{W}_{j,t}^{\gamma})^\top | \mathcal{F}) = \sum_{u,v,k,l=1}^{d} \partial_{uv} \gamma_{gi} \left( \int_0^t c_s ds \right) \left( \int_0^t (c_{uk,s} c_{vl,s} + c_{ul,s} c_{vk,s}) ds \right) \partial_{kl} \gamma_{gj} \left( \int_0^t c_s ds \right), \]

\[ \partial_{kl} \gamma_{gi} (A) = (\lambda_g (A) I - A)_{ik}^{+} \gamma_{gl} (A). \]

The corresponding principal component is then defined as \( \gamma_g^T (X_t - X_0) \). However, this eigenvalue and the principal component do not satisfy the fundamental relationship between the eigenvalue and the variance of the principal component:

\[ \lambda_g \left( \int_0^t c_s ds \right) \notin [\gamma_g^T (X_t - X_0), \gamma_g^T (X_t - X_0)]^c, \tag{20} \]

which is a desired property of any sensible PCA procedure. This fact alone makes this second procedure much less useful than the one we proposed above, based on integrated eigenvalues of the spot covariance. Another feature that distinguishes this second “PCA” procedure with both the classical one and our realized PCA is that the asymptotic covariance matrix of eigenvalues of \( \int_0^t c_s ds \) is no longer diagonal.

An analogy of the relationship between the two PCA procedures is that between integrated quarticity \( t \int_0^t \sigma^4_s ds \) and the squared integrated volatility \( (\int_0^t \sigma_s^2 ds)^2 \). If the covariance matrix is constant, the two procedures are equivalent but not in general.

### 4 Simulation Evidence

We now investigate the small sample performance of the estimators in conditions that approximate the empirical setting of PCA for large stock portfolios. In order to generate data with a common factor structure, we simulate a factor model in which a vector of log-stock prices \( X \) follows the continuous-time dynamics

\[ dX_t = \beta_t dF_t + dZ_t, \tag{21} \]

where \( F \) is a collection of unknown factors, \( \beta \) is the matrix of factor loadings, and \( Z \) is a vector of idiosyncratic components, which is orthogonal to \( F \). This model is a special case of the general semimartingale model (2).

By construction, the continuous part of the quadratic variation has the following local structure:

\[ [dX, dX]_t^c = \beta_t [dF, dF]_t^c \beta_t^\top + [dZ, dZ]_t^c. \]

When the idiosyncratic components have smaller magnitude, the dominating eigenvalues of \( X \) are close to the non-zero eigenvalues of \( \beta_t [dF, dF]_t^c \beta_t^\top \), by Weyl’s inequality. As a result, we should
expect three large eigenvalues from the scree plot. Our goal here is to conduct nonparametric PCA rather than to estimate a parametric factor model, for which we need to resort to a large panel of \( X \), whose dimension increases with the sample size, see, e.g., Bai and Ng (2002), and so (21) is only employed as the data-generating process on which PCA is employed without any knowledge of the underlying factor structure.

Specifically, we simulate

\[
\begin{align*}
    dX_{i,t} &= \sum_{j=1}^{r} \beta_{ij,t} dF_{j,t} + dZ_{i,t}, \\
    dF_{j,t} &= \mu_{j,t} dt + \sigma_{j,t} dW_{j,t} + dJ_{j,t}, \\
    dZ_{i,t} &= \gamma_{i,t} dB_{i,t} + dJ'_{i,t},
\end{align*}
\]

where \( i = 1, 2, \ldots, d \), and \( j = 1, 2, \ldots, r \). In the simulations, one of the \( F_s \) plays the role of the market factor, so that its associated \( \beta \)s are positive. The correlation matrix of \( dW \) is denoted as \( \rho^F \). We allow for time-varying \( \sigma_{j,t} \), \( \gamma_{i,t} \), and \( \beta_{ij,t} \), which evolve according to the following system of equations:

\[
\begin{align*}
    d\sigma_{j,t}^2 &= \kappa_j (\theta_j - \sigma_{j,t}^2) dt + \eta_j \sigma_{j,t} d\tilde{W}_{j,t} + dJ_{j,t}^2, \\
    d\gamma_{i,t}^2 &= \kappa (\theta - \gamma_{i,t}^2) dt + \eta \gamma_{i,t} dB_t, \\
    d\beta_{ij,t} &= \begin{cases} 
        \tilde{\kappa}_j (\tilde{\theta}_{ij} - \beta_{ij,t}) dt + \tilde{\xi}_j \sqrt{\beta_{ij,t}} d\tilde{B}_{ij,t} & \text{if the } j \text{th factor is the “market”}, \\
        \tilde{\kappa}_j (\tilde{\theta}_{ij} - \beta_{ij,t}) dt + \tilde{\xi}_j dB_{ij,t} & \text{otherwise},
    \end{cases}
\end{align*}
\]

where the correlation between \( dW_j \) and \( d\tilde{W}_j \) is \( \rho^F \), \( \{J_{1}^F \}_{1 \leq j \leq r} \) and \( \{J_{1}^Z \}_{1 \leq i \leq d} \) are driven by two Poisson Processes with arrival rates \( \lambda^F \) and \( \lambda^Z \), respectively. Their jump sizes follow double exponential distributions with means denoted by \( \mu_{+}^F \), \( \mu_{-}^F \), \( \mu_{+}^Z \), and \( \mu_{-}^Z \), respectively. \( \{J_{1}^{\sigma^2} \}_{1 \leq j \leq r} \) co-jumps with \( J^F \), and their jump sizes follow exponential distributions with the mean equal to \( \mu^2 \). All the model parameters are given in Table 1. They are chosen to be realistic given the empirical characteristics of the cross-section of stock returns and their volatilities.

Anticipating our empirical application to the S&P 100 Index constituents, we simulate intraday returns of \( d = 100 \) stocks at various frequencies and with horizons \( T \) spanning 1 week to 1 month. When \( r = 3 \), the model implies 3 distinct eigenvalues reflecting the local factor structure of the simulated data. The remaining population eigenvalues are identical, due to idiosyncratic variations. Throughout, we fix \( k_n \) to be the closest divisors of \( t/\Delta_n \) to \( \theta \Delta_n^{-1/2} \sqrt{\log(d)} \), with \( \theta = 0.5 \) and \( d \) is the dimension of \( X \). Our choice of \( \sqrt{\log(d)} \) is motivated from the literature on high-dimensional covariance matrix estimation, although our asymptotic design does not take into account an increasing dimensionality. Other choices of \( \theta \), e.g., 0.05 - 0.5, or functions of \( d \), e.g., \( \sqrt{d\log d} \), deliver the same results. To truncate off jumps from spot covariance estimates, we adopt the usual procedure in the literature, i.e., choosing \( u_{i,n} = 3\left( \int_{0}^{t} c_{ii,s} ds / t \right)^{0.5} \Delta_n^{0.47} \), for \( 1 \leq i \leq d \), where \( \int_{0}^{t} c_{ii,s} ds \) can be estimated by, for instance, bipower variations.

We then apply the realized PCA procedure. We first examine the curse of dimensionality –
how increasing number of stocks affects the estimation – and how the sampling frequency affects the estimation. In light of Corollary 1, we estimate 3 simple integrated eigenvalues as well as the average of the remaining identical eigenvalues, denoted as $\int_0^t \lambda_i ds$, $i = 1, 2, 3$, and 4. We report the mean and standard errors of the estimates as well as the root-mean-square errors of the standardized estimates with $d = 5, 10, 15, 20, 30, 50$, and 100 stocks, respectively, using returns sampled every $\Delta_n = 5$ seconds, 1 minute and 5 minutes over one week and one month horizons.

The results, as shown from Tables 2 - 5, suggest that, as expected, the estimation is more difficult as the dimensionality increases, but the large amount of high frequency data and in-fill asymptotic techniques deliver very satisfactory finite sample approximations. The eigenvalues are accurately recovered. The repeated eigenvalues are estimated with smaller biases and standard errors, due to the extra averaging taken at the estimation stage. In Tables 6 - 7, we provide estimates for the first eigenvectors. Similar to the estimation for eigenvalues, the estimates are very accurate, even with 100 stocks.

To further verify the accuracy of the asymptotic distribution in small samples as the sample size increases, we provide in Figure 1 histograms of the standard estimates of integrated eigenvalues using 30 stocks with 5-second returns. Finally, we examine the finite sample accuracy of the integrated simple eigenvalues, in the more challenging scenario with 100 stocks sampled every minute, which matches the setup of the empirical analysis we will conduct below. To verify that the detection of 3 eigenvalues is not an artifact, we vary the number of common factors in the data generating process from $r = 2$ to 4, by removing the third factor or adding another factor that shares the same parameters as the third factor. The histograms are provided in Figure 2. The results collectively show that the finite sample performance of the method is quite good even for a 100-stock portfolio with one-minute returns and one-week horizon.

5 High-Frequency Principal Components in the S&P 100 Stocks

Given the encouraging Monte Carlo results, we now conduct PCA on intraday returns of S&P 100 Index (OEX) constituents. We collect stock prices over the 2003 - 2012 period from the Trade and Quote (TAQ) database of the New York Stock Exchange (NYSE).\footnote{While the constituents of the OEX Index change over time, we keep track of the changes to ensure that our choice of stocks is always in line with the index constituents.} Due to entry and exit from the index, there are in total 154 tickers over this 10-year period. We conduct PCA on a weekly basis. One of the key advantages of the large amount of high frequency data is that we are effectively able to create a ten-year long time series of eigenvalues and principal components at the weekly frequency by collating the results obtained over each week in the sample.

After removing those weeks during which the index constituents are switching, we are left with 482 weeks. To clean the data, for each ticker on each trading day, we only keep the intraday prices
from the single exchange which has the largest number of transaction records. We provide in Figure 3 the quantiles of the number of transactions between 9:35 a.m. EST and 3:55 p.m. EST, across 100 stocks each day. These stocks have excellent liquidity over the sampling period. We thereby employ 1-minute subsamples for the most liquid 90 stocks in the index (for simplicity, we still refer to this 90-stock subsample as the S&P 100) in order to address any potential concerns regarding microstructure noise and asynchronous trading.\(^5\) These stocks trade multiple times per sampling interval and we compute returns using the latest prices recorded within each sampling interval. Overnight returns are removed so that there is no concern of price changes due to dividend distributions or stock splits. The 158 time series of cumulative returns are plotted in Figure 4.

### 5.1 Scree Plot

We report the time series average of the eigenvalue estimates against their order in Figure 5, a plot known in classical PCA analysis as the “scree plot” (see e.g. Jolliffe (2002)). This plot classically constitutes the main graphical aid employed to determine the appropriate number of components in empirical applications. The graph reveals that there are a handful of eigenvalue estimates, three, which are separated from the rest. The fact that three factors on average explain the cross-sectional variation of stock returns is surprisingly consistent with the well-established low frequency Fama-French common factor analysis, despite the large differences in methods, time periods, sample frequencies and length of observation.

We also plot the estimates of the percentage variation explained by the first three integrated eigenvalues in Figure 6, along with the average variation explained by the remaining eigenvalues. There are substantial time variation of the first three eigenvalues, all of which are at peak around the recent financial crisis, indicating an increased level of comovement, which in this context is the definition of systemic risk. The idiosyncratic factors become relatively less important and even more dominated by the common factors during the crisis.

The first eigenvalue accounts for on average 30-40% of the total variations of the 90 constituents, capturing the extent of the market variation in the sample. The second and third components together account for an additional 15%-20% of the total variation. Therefore, there exists significant amount of remaining idiosyncratic variation, beyond what can be explained by a three-common-factor model. The average variation explained by the remaining 87 components are around 0.4%-0.6%, which corresponds to the idiosyncratic contributions relative to the first three principal components.

Next, we compare the cumulative returns of the first three cumulative principal components.

\(^5\) Estimators that are robust to both noise and asynchronicity are available for integrated covariance estimation in Aït-Sahalia, Fan, and Xiu (2010), Christensen, Kinnebrock, and Podolskij (2010), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011), and Shephard and Xiu (2012). Extending the current method to a noisy and asynchronous setting is theoretically interesting but not empirically necessary for the present paper; it is left for future work.
The time series plots are shown in Figure 7. The empirical correlations among the three components are -0.020, 0.005, and -0.074, respectively, which agrees with the design that the components are orthogonal. The first principal component accounts for most of the common variation, and it shares the time series features of the overall market return. This is further reinforced by the fact that the loadings of all the stocks on the first principal component, although time-varying, remain positive throughout the sample period, which is not the case for the additional principal components. It is worth pointing out however that these principal components are not constrained to be portfolio returns, as their weights at each point in time are nowhere constrained to add up to one. This means that the first principal component cannot be taken directly to be the market portfolio, or more generally identified with additional Fama-French or additional mimicking portfolios.

5.2 Biplots

Although PCA is by design not suited to identifying the underlying economic factors corresponding to the principal components, the time series evidence in Figure 7 suggests that the first principal component captures the features of the overall market return, as already noted. Can we extract more information about the economic nature of the principal components? For this purpose, we can use our results to produce biplots (see Gabriel (1971)) and see which assets line up together in the basis of the principal components. A biplot gives the relative position of the $d$ variables on a plot where two principal components are the axes. The length of a vector from the origin pointing towards the position of a variable on the biplot tends to reflect the magnitude of the variation of the variable. Moreover, vectors associated with similar variables tend to point towards the same direction, hence a biplot can be used for classification of the assets. While it is possible to compute biplots for the spot eigenvectors, the biplots for the integrated eigenvectors are more stable.

We report two representative snapshots of the biplots of the first two integrated eigenvectors in Figures 8 (March 7-11, 2005, preceding the crisis) and 9 (March 10-14, 2008, during the crisis, although not at its most extreme), in order to interpret the second principal component. We identify differently in the figures the vectors that correspond to financial stocks (identified using the Global Industrial Classification Standard (GICS) codes). The Federal Reserve made several important announcement during the week of March 10 -14, 2008 to address heightened liquidity pressures of the financial system, including the creation of the Term Securities Lending Facility, and the approval of a financing arrangement between J.P. Morgan Chase and Bear Stearns.

We find that the financial sector is clearly separated from the rest of the stocks during March 10-14, 2008, in sharp contrast with the biplot for the week of March 7 - 11, 2005, during which no such separation occurs. This suggests that PCA based on low-frequency data is likely to overlook these patterns. Interestingly, we can also see from the March 2008 biplot that the Lehman Brothers (LEH) stands out, having the largest loadings on both components. The corresponding scree plots
for these two weeks in Figures 10 and 11 both suggest the presence of at least three factors, but the eigenvalues in March 10-14, 2008 are much larger in magnitude, consistently with a higher systematic component to asset returns. For each week, the figures report 95% confidence intervals for the first three eigenvalues computed based on the distribution given in Corollary 1.

6 Conclusions

This paper develops the tools necessary to implement PCA at high frequency, constructing and estimating realized eigenvalues, eigenvectors and principal components. This development complements the classical PCA theory in a number of ways. Compared to its low frequency counterpart, PCA becomes feasible over short windows of observation (of the order of one week), relatively large dimensions (90 in our application) and further are free from the need to impose strong parametric assumptions on the distribution of the data, applying instead to a broad class of semimartingales. The estimators perform well in simulations and reveal that the joint dynamics of the S&P 100 stocks at high frequency are well explained by a three-factor model, a result that is broadly consistent with the Fama-French factor model at low frequency, surprisingly so given the large differences in time scale, sampling and returns horizon.

This paper represents a necessary first step to bring PCA tools to a high frequency setting. Although not empirically relevant in the context of the analysis above of a portfolio of highly liquid stocks at the one-minute frequency, the next steps in the development of the theory will likely include the incorporation of noise-robust and asynchronicity-robust covariance estimation methods. We hope to pursue these extensions in future work.
References


Appendix A  Figures and Tables

<table>
<thead>
<tr>
<th>j</th>
<th>$\kappa_j$</th>
<th>$\theta_j$</th>
<th>$\eta_j$</th>
<th>$\rho_j$</th>
<th>$\mu_j$</th>
<th>$\kappa_j$</th>
<th>$\theta_{i,j}$</th>
<th>$\xi_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>0.05</td>
<td>0.3</td>
<td>-0.6</td>
<td>0.05</td>
<td>1</td>
<td>$U[0.25, 1.75]$</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.04</td>
<td>0.4</td>
<td>-0.4</td>
<td>0.03</td>
<td>2</td>
<td>$\mathcal{N}(0, 0.5^2)$</td>
<td>0.6</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0.03</td>
<td>0.3</td>
<td>-0.25</td>
<td>0.02</td>
<td>3</td>
<td>$\mathcal{N}(0, 0.5^2)$</td>
<td>0.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\lambda^F$</th>
<th>$\mu^F_{+/-}$</th>
<th>$\lambda^Z$</th>
<th>$\mu^Z_{+/-}$</th>
<th>$\mu^\sigma^2$</th>
<th>$\rho^F_{12}$</th>
<th>$\rho^F_{13}$</th>
<th>$\rho^F_{23}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1/t$</td>
<td>$4\sqrt{\Delta}$</td>
<td>$2/t$</td>
<td>$6\sqrt{\Delta}$</td>
<td>$\sqrt{\Delta}$</td>
<td>0.05</td>
<td>0.1</td>
<td>0.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\kappa$</th>
<th>$\theta$</th>
<th>$\eta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$4$</td>
<td>0.3</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Table 1: Parameters in Monte Carlo Simulations

Note: In this table, we report the parameter values used in the simulations. The constant matrix $\hat{\theta}_{i,j}$ is generated randomly from the described distribution, and is fixed throughout all replications. The dimension of $X_t$ is 100, whereas the dimension of $F_t$ is 3. $\Delta$ is the sampling frequency, and $t$ is the length of the time window. The number of Monte Carlo replications is 1,000.
<table>
<thead>
<tr>
<th># Stocks</th>
<th>1 Week, 5 Seconds</th>
<th>1 Week, 1 Minute</th>
<th>1 Week, 5 Minutes</th>
<th>1 Month, 5 Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True</td>
<td>Bias</td>
<td>Stdev</td>
<td>True</td>
</tr>
<tr>
<td>5</td>
<td>0.4686</td>
<td>-0.0001</td>
<td>0.0075</td>
<td>0.4703</td>
</tr>
<tr>
<td>10</td>
<td>0.6489</td>
<td>0.0001</td>
<td>0.0110</td>
<td>0.6552</td>
</tr>
<tr>
<td>15</td>
<td>0.8927</td>
<td>-0.0004</td>
<td>0.0149</td>
<td>0.8975</td>
</tr>
<tr>
<td>20</td>
<td>1.3044</td>
<td>0.0003</td>
<td>0.0225</td>
<td>1.3148</td>
</tr>
<tr>
<td>30</td>
<td>2.1003</td>
<td>-0.0002</td>
<td>0.0356</td>
<td>2.1134</td>
</tr>
<tr>
<td>50</td>
<td>2.9863</td>
<td>0.0002</td>
<td>0.0514</td>
<td>3.0104</td>
</tr>
<tr>
<td>100</td>
<td>6.6270</td>
<td>-0.0004</td>
<td>0.1141</td>
<td>6.6732</td>
</tr>
<tr>
<td></td>
<td>0.4879</td>
<td>0.0107</td>
<td>0.0452</td>
<td>0.5642</td>
</tr>
<tr>
<td>10</td>
<td>0.6839</td>
<td>0.0084</td>
<td>0.0574</td>
<td>0.7143</td>
</tr>
<tr>
<td>15</td>
<td>0.9397</td>
<td>0.0128</td>
<td>0.0724</td>
<td>1.0167</td>
</tr>
<tr>
<td>20</td>
<td>1.3765</td>
<td>0.0130</td>
<td>0.1076</td>
<td>1.3882</td>
</tr>
<tr>
<td>30</td>
<td>2.2157</td>
<td>0.0210</td>
<td>0.1670</td>
<td>2.2383</td>
</tr>
<tr>
<td>50</td>
<td>3.1554</td>
<td>0.0267</td>
<td>0.2410</td>
<td>3.1518</td>
</tr>
<tr>
<td>100</td>
<td>7.0000</td>
<td>0.0552</td>
<td>0.5314</td>
<td>6.9632</td>
</tr>
</tbody>
</table>

Table 2: Simulation Results: First Eigenvalue Estimation

Note: In this table, we report the summary statistics of 1,000 Monte Carlo simulations for estimating the first integrated eigenvalue. Column “True” corresponds to the average of the true integrated eigenvalue; Column “Bias” corresponds to the mean of the estimation error; Column “Stdev” is the standard deviation of the estimation error.
Table 3: Simulation Results: Second Eigenvalue Estimation

Note: In this table, we report the summary statistics of 1000 Monte Carlo simulations for estimating the second integrated eigenvalue. Column “True” corresponds to the average of the true integrated eigenvalue; Column “Bias” corresponds to the mean of the estimation error; Column “Stdev” is the standard deviation of the estimation error.
<table>
<thead>
<tr>
<th># Stocks</th>
<th>1 Week, 5 Seconds</th>
<th>1 Week, 1 Minute</th>
<th>1 Week, 5 Minutes</th>
<th>1 Month, 5 Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.1338 0.0001 0.0022</td>
<td>0.1345 0.0003 0.0044</td>
<td>0.1364 0.0041 0.0124</td>
<td>0.1404 0.0006 0.0054</td>
</tr>
<tr>
<td>10</td>
<td>0.2132 0.0001 0.0035</td>
<td>0.2149 0.0002 0.0070</td>
<td>0.2199 0.0033 0.0183</td>
<td>0.2367 -0.0001 0.0091</td>
</tr>
<tr>
<td>15</td>
<td>0.2941 0.0002 0.0046</td>
<td>0.2954 0.0002 0.0093</td>
<td>0.3018 0.0056 0.0264</td>
<td>0.3396 0.0009 0.0131</td>
</tr>
<tr>
<td>20</td>
<td>0.3212 0.0001 0.0052</td>
<td>0.3242 0.0000 0.0105</td>
<td>0.3242 0.0000 0.0105</td>
<td>0.3546 0.0003 0.0132</td>
</tr>
<tr>
<td>30</td>
<td>0.4825 0.0005 0.0078</td>
<td>0.4853 0.0001 0.0153</td>
<td>0.4971 0.0055 0.0409</td>
<td>0.5686 0.0006 0.0211</td>
</tr>
<tr>
<td>50</td>
<td>0.6943 0.0001 0.0113</td>
<td>0.7016 -0.0005 0.0226</td>
<td>0.7016 -0.0005 0.0226</td>
<td>0.8051 -0.0010 0.0306</td>
</tr>
<tr>
<td>100</td>
<td>1.3808 0.0004 0.0221</td>
<td>1.3935 -0.0013 0.0438</td>
<td>1.3808 0.0004 0.0221</td>
<td>1.4302 -0.0016 0.1119</td>
</tr>
</tbody>
</table>

Table 4: Simulation Results: Third Eigenvalue Estimation

Note: In this table, we report the summary statistics of 1,000 Monte Carlo simulations for estimating the third integrated eigenvalue. Column “True” corresponds to the average of the true integrated eigenvalue; Column “Bias” corresponds to the mean of the estimation error; Column “Stdev” is the standard deviation of the estimation error.
<table>
<thead>
<tr>
<th># Stocks</th>
<th>1 Week, 5 Seconds</th>
<th>1 Week, 1 Minute</th>
<th>1 Week, 5 Minutes</th>
<th>1 Month, 5 Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True</td>
<td>Bias</td>
<td>Stdev</td>
<td>True</td>
</tr>
<tr>
<td>5</td>
<td>0.0596</td>
<td>0.0002</td>
<td>0.0007</td>
<td>0.0597</td>
</tr>
<tr>
<td>10</td>
<td>0.0596</td>
<td>0.0001</td>
<td>0.0004</td>
<td>0.0597</td>
</tr>
<tr>
<td>15</td>
<td>0.0596</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0597</td>
</tr>
<tr>
<td>20</td>
<td>0.0596</td>
<td>0.0001</td>
<td>0.0002</td>
<td>0.0597</td>
</tr>
<tr>
<td>30</td>
<td>0.0596</td>
<td>0.0001</td>
<td>0.0002</td>
<td>0.0597</td>
</tr>
<tr>
<td>50</td>
<td>0.0596</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0597</td>
</tr>
<tr>
<td>100</td>
<td>0.0596</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0597</td>
</tr>
</tbody>
</table>

Table 5: Simulation Results: Repeated Eigenvalue Estimation, Fourth and Beyond

Note: In this table, we report the summary statistics of 1,000 Monte Carlo simulations for estimating the repeated integrated eigenvalues. Column “True” corresponds to the average of the true integrated eigenvalue; Column “Bias” corresponds to the mean of the estimation error; Column “Stdev” is the standard deviation of the estimation error.
<table>
<thead>
<tr>
<th># Stocks</th>
<th>True</th>
<th>Bias</th>
<th>Stdev</th>
<th>True</th>
<th>Bias</th>
<th>Stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1 Week, 5 Seconds</td>
<td>0.2457</td>
<td>-0.0002</td>
<td>0.0173</td>
<td>0.2475</td>
<td>0.0074</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.8987</td>
<td>0.0028</td>
<td>0.0261</td>
<td>0.8963</td>
<td>0.0174</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0566</td>
<td>0.0013</td>
<td>0.0235</td>
<td>0.0536</td>
<td>-0.0018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0633</td>
<td>0.0001</td>
<td>0.0119</td>
<td>0.0639</td>
<td>0.0037</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3110</td>
<td>0.0017</td>
<td>0.0202</td>
<td>0.3086</td>
<td>0.0042</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>0.0506</td>
<td>0.0004</td>
<td>0.0049</td>
<td>0.0495</td>
<td>0.0029</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3444</td>
<td>0.0006</td>
<td>0.0087</td>
<td>0.3452</td>
<td>0.0080</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4632</td>
<td>0.0011</td>
<td>0.0129</td>
<td>0.4619</td>
<td>0.0121</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1422</td>
<td>0.0008</td>
<td>0.0137</td>
<td>0.1422</td>
<td>0.0094</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4166</td>
<td>0.0004</td>
<td>0.0220</td>
<td>0.4164</td>
<td>0.0045</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0864</td>
<td>0.0008</td>
<td>0.0158</td>
<td>0.0863</td>
<td>0.0089</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3460</td>
<td>0.0008</td>
<td>0.0088</td>
<td>0.3440</td>
<td>0.0101</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1268</td>
<td>0.0005</td>
<td>0.0076</td>
<td>0.1259</td>
<td>0.0066</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3409</td>
<td>0.0005</td>
<td>0.0174</td>
<td>0.3415</td>
<td>0.0046</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4262</td>
<td>0.0007</td>
<td>0.0112</td>
<td>0.4271</td>
<td>0.0099</td>
</tr>
</tbody>
</table>

Table 6: Simulation Results: Eigenvector Estimation

Note: In this table, we report the summary statistics of 1,000 Monte Carlo simulations for estimating the integrated eigenvector associated with the first eigenvalue. Column “True” corresponds to the average of the true integrated vector; Column “Bias” corresponds to the mean of the estimation error; Column “Stdev” is the standard deviation of the estimation error. The setup for these simulations is identical to that of Tables 2-5.
Table 7: Simulation Results: Eigenvector Estimation

Note: In this table, we report the summary statistics of 1,000 Monte Carlo simulations for estimating the integrated eigenvectors associated with the first eigenvalues. The column “True” corresponds to the average of the true integrated vector; “Bias” corresponds to the mean of the estimation error; “Stdev” is the standard deviation of the estimation error. The setup for these simulations is identical to that of Tables 2-5. To save space, we do not report the eigenvectors in dimensions larger than 30.
Figure 1: Finite Sample Distribution of the Standardized Statistics.

Note: In this figure, we report the histograms of the 1,000 simulation results for estimating the first four integrated eigenvalues using 5-second returns for 30 stocks over one week. The purpose of this figure is to validate the asymptotic theory, hence the use of a short 5-second sampling interval. The smallest 27 eigenvalues are identical so that the fourth eigenvalue is repeated. The solid lines plot the standard normal density; the dashed histograms report the distribution of the estimates before bias correction; the solid histograms report the distribution of the estimates after bias correction is applied and is to be compared to the asymptotic standard normal. Because the fourth eigenvalue is small, the dashed histogram on the fourth subplot is out of the x-axis range, to the right.
Figure 2: Finite Sample Distribution of the Standardized Statistics.

Note: In this figure, we report the histograms of the integrated simple eigenvalues using weekly one-minute returns of 100 stocks, a setting that matches that of our empirical analysis below. Columns 1, 2, and 3 report the results with $r = 2, 3, \text{ and } 4$ common factors in the data generating process, respectively. The number of Monte Carlo replications is 1,000.
Figure 3: The Liquidity of S&P 100 Index Components

Note: In this figure, we provide quartiles of the average time between successive transactions, or sampling frequencies, for the 100 stocks S&P 100 Index constituents between 2003 and 2012. We exclude the first and the last 5 minutes from the 6.5 regular trading hours in the calculation, during which trading intensities are unusually high. The y-axis reports the corresponding average sampling frequencies in seconds, computed from the high frequency transactions in the sample. From these transactions, we construct a synchronous one-minute sample using the latest tick recorded within that minute. The 10 least liquid stocks in the index are included in this figure but excluded from the empirical analysis that follows.
Note: In this figure, we plot the time series of cumulative daily open-to-close returns of 158 S&P 100 Index Components from 2003 to 2012. The thick black solid line plots the cumulative daily open-to-close S&P 100 Index returns. All overnight returns are excluded. Companies that exited the index during the sample period, including bankruptcies, are represented by a time series truncated at the time of delisting.
Figure 5: The Scree Graph

Note: In this figure, we report the time series average over the entire sample of the estimated eigenvalues against their order, or “scree graph”. These integrated eigenvalues are estimated assuming that all of them are simple.
Figure 6: Time Series of Realized Eigenvalues

Note: In this figure, we plot the time series of the three largest realized eigenvalues, representing the percentage variations explained by the corresponding principal components, using weekly 1-min returns of 90 most liquid S&P 100 Index constituents from 2003 to 2012. For comparison, we also plot the average of the remaining 87 eigenvalues.
Figure 7: Time Series of Cumulative Realized Principal Components

Note: In this figure, we compare the cumulative returns of the first three principal components, using weekly 1-minute returns of the S&P 100 Index constituents from 2003 to 2012.
Figure 8: Biplot of the Integrated Eigenvector: Pre-Crisis Week

Note: This biplot shows the integrated eigenvector estimated during the week March 7-11, 2005. Each dot and the vector joining it and the origin represents a firm. The x and y coordinates of a firm denote its average loading on the first and second principal components, respectively. The dots and lines associated with firms in the financial sector are colored in red and dashed, while non-financial firms are colored in blue and drawn as solid lines. The figure shows little difference between the two groups during that week.
Note: This biplot shows the integrated eigenvector estimated during the week March 10-14, 2008. Each dot and the vector joining it and the origin represents a firm. The x and y coordinates of a firm denote its average loading on the first and second principal components, respectively. The dots and lines associated with firms in the financial sector are colored in red and dashed, with the following singled out: American International Group (AIG), Bank of America (BOA), Citigroup (C), Goldman Sachs (GS), J.P. Morgan Chase (JPM), and Lehman Brothers (LEH). Non-financial firms are colored in blue and drawn as solid lines. The figure shows a sharp distinction between the two groups’ dependence on the second principal component during that week.
Figure 10: Scree Plot of the Integrated Eigenvalues: Pre-Crisis Week

Note: This figure contains a scree plot of the integrated eigenvalues estimated during the week March 7-11, 2005. The blue solid lines provide 95% confidence interval for the first three eigenvalues computed based on the results of Corollary 1.
Figure 11: Scree Plot of the Integrated Eigenvalues: Crisis Week

Note: This figure contains a scree plot of the integrated eigenvalues estimated during the week March 10-14, 2008. The blue solid lines provide 95% confidence interval for the first three eigenvalues computed based on the results of Corollary 1.
Appendix B  Mathematical Proofs

Appendix B.1  Proof of Lemma 1

Proof. By the classical Weyl inequalities in e.g., Horn and Johnson (2013) and Tao (2012), we have \(|\lambda_j(A + \epsilon) - \lambda_j(A)| \leq K \|\epsilon\|\), for all \(j = 1, 2, \ldots, d\), where \(A, \epsilon \in \mathcal{M}^+_d\), and \(K\) is some constant. This establishes the Lipchitz property.

Appendix B.2  Proof of Lemma 2

Proof. It is straightforward to show (by the implicit function theorem, see Magnus and Neudecker (1999) Theorem 8.7) that any simple eigenvalue and its corresponding eigenvector, written as functions of \(A\), \(\lambda_g(A)\) and \(\gamma_g(A)\), are \(C^\infty\). To calculate their derivatives, note that \(A\gamma_g = \lambda_g \gamma_g\), hence we have \((\partial_{jk} A)\gamma_g + A(\partial_{jk} \gamma_g) = (\partial_{jk} \lambda_g)\gamma_g + \lambda_g(\partial_{jk} \gamma_g)\). Pre-multiplying \(\gamma_g^T\) on both sides yields \(\partial_{jk} \lambda_g = \gamma_g^T(\partial_{jk} A)\gamma_g = \gamma_{pj} \gamma_{gk}\). Rewrite it into \((\lambda_g I - A)\partial_{jk} \gamma_g = (\partial_{jk} A)\gamma_g - (\partial_{jk} \lambda_g)\gamma_g\), which leads to \((\lambda_g I - A)^+ (\lambda_g I - A)\partial_{jk} \gamma_g = (\lambda_g I - A)^+(\partial_{jk} A)\gamma_g\). As a result, \(\partial_{jk} \gamma_g = (\lambda_g I - A)^+ J_{jk} \gamma_g\).

In the case when all eigenvalues are simple, by direct calculation we have

\[
\partial_{jk} \gamma_{gh} = \sum_{p \neq g} \frac{1}{\lambda_g - \lambda_p} \gamma_{ph} \gamma_{pj} \gamma_{gk},
\]

where we use the fact that \((\gamma_1^T, \gamma_2^T, \ldots, \gamma_d^T)^T A(\gamma_1, \gamma_2, \ldots, \gamma_d) = \text{Diag}(\lambda(A))\). Further,

\[
\partial^2_{jk,lm} \gamma_{gh} = -\sum_{p \neq g} \frac{1}{(\lambda_g - \lambda_p)^2} (\partial_{lm} \lambda_g - \partial_{lm} \lambda_p) \gamma_{ph} \gamma_{pj} \gamma_{gk} + \sum_{p \neq g} \frac{1}{\lambda_g - \lambda_p} \partial_{lm} \left( \gamma_{ph} \gamma_{pj} \gamma_{gk} \right)
\]

\[
= -\sum_{p \neq g} \frac{1}{(\lambda_g - \lambda_p)^2} \left( \gamma_{ql} \gamma_{gm} \gamma_{ph} \gamma_{pj} \gamma_{gk} - \gamma_{ph} \gamma_{pj} \gamma_{gk} \right)
\]

\[+ \sum_{p \neq g} \sum_{q \neq p} \frac{1}{(\lambda_g - \lambda_p)(\lambda_p - \lambda_q)} \gamma_{ql} \gamma_{pm} \gamma_{ph} \gamma_{pj} \gamma_{gk}
\]

\[+ \sum_{p \neq g} \sum_{q \neq p} \frac{1}{(\lambda_g - \lambda_p)(\lambda_p - \lambda_q)} \gamma_{ql} \gamma_{pm} \gamma_{qj} \gamma_{ph} \gamma_{gk}
\]

\[+ \sum_{p \neq g} \sum_{q \neq p} \frac{1}{(\lambda_g - \lambda_p)(\lambda_p - \lambda_q)} \gamma_{qk} \gamma_{ql} \gamma_{gm} \gamma_{ph} \gamma_{pj},
\]

which concludes the proof.

Appendix B.3  Proof of Lemma 3

Proof. The proof is by induction. Consider, at first the following optimization problem:

\[
\max_{\gamma_s} \int_0^u \gamma_s^T c_s \gamma_s ds, \ \text{s.t.} \ \gamma_s^T \gamma_s = 1, \ \ 0 \leq s \leq u \leq t.
\]

Using a sequence of Lagrange multipliers \(\lambda_s\), the problem can be written as solving

\[
c_s \gamma_s = \lambda_s \gamma_s, \ \text{and} \ \gamma_s^T \gamma_s = 1, \ \text{for any} \ 0 \leq s \leq t.
\]
Hence, the original problem is translated into eigenanalysis.

Suppose the eigenvalues of $c_s$ are ordered as in $\lambda_{1,s} \geq \lambda_{2,s} \geq \ldots \geq \lambda_{d,s}$. Note that $\gamma_s^T c_s \gamma_s = \lambda_s$, so that $\lambda_s = \lambda_{1,s}$, and $\gamma_s = \gamma_{1,s}$ is one of the corresponding eigenvectors (if $\lambda_{1,s}$ is not unique), and the maximal variation is $\int_0^t \lambda_{1,s} ds$.

Suppose that we have found $\gamma_{1,s}, \ldots, \gamma_{k,s}$, for $1 \leq k < d$ and $0 \leq s \leq t$, the $(k+1)$th principal component is defined by solving the following problem:

$$\max_{\gamma_s} \int_0^u \gamma_s^T c_s \gamma_s ds, \quad \text{s.t.} \quad \gamma_s^T \gamma_s = 1, \quad \text{and} \quad \gamma_s^T c_s \gamma_s = 0, \quad 1 \leq j \leq k, \quad 0 \leq u \leq t.$$  

Using similar technique of Lagrange multipliers, $\lambda_s$, and $\nu_{1,s}, \ldots, \nu_{k,s}$, we find

$$c_s \gamma_s = \lambda_s \gamma_s + \sum_{j=1}^k \nu_{j,s} c_s \gamma_{j,s}.$$  

Multiplying on the left $\gamma_{l,s}$, for some $1 \leq l \leq k$, we can show that $\nu_{l,s} c_s \gamma_{l,s} = 0$. Indeed,

$$0 = \lambda_{l,s} \gamma_{l,s}^T \gamma_s = \gamma_{l,s}^T c_s \gamma_s = \gamma_{l,s}^T \lambda_s \gamma_s + \sum_{j=1}^k \nu_{j,s} \gamma_{l,s}^T c_s \gamma_{j,s} = \nu_{l,s} \lambda_{l,s}.$$  

Therefore, since $l$ is an arbitrary number between $1$ and $k$, we have $c_s \gamma_s = \lambda_s \gamma_s$. Hence, $\lambda_s = \lambda_{k+1,s}$, $\gamma_s = \gamma_{k+1,s}$ is one of the eigenvectors associated with the eigenvalue $\lambda_{k+1,s}$. This establishes the first part of the theorem.

For any càdlàg and adapted process $\gamma_s$,

$$\left[ \int_0^u \gamma_s^T dX_s, \int_0^u \gamma_s^T dX_s \right] = \int_0^t\gamma_s^T c_s \gamma_s ds.$$  

Hence the statement follows from the $g$th-step optimization problem. Note that the validity of the integrals above is warranted by the continuity of $\lambda$ given by Lemma 1.

\section*{Appendix B.4 Proof of Lemma 4}

\textit{Proof.} The first statement of the proof follows by immediate calculations from Theorem 1.1 in Lewis (1996b) and Theorem 3.3 in Lewis and Sendov (2001). The second statement is discussed and proved in, e.g., Ball (1984), Sylvester (1985), and Silhavý (2000). Finally, the last statement on convexity is proved in Davis (1957) and Lewis (1996a).

\section*{Appendix B.5 Proof of Lemma 5}

\textit{Proof.} Obviously, for any $1 \leq g_1 < g_2 < \ldots < g_r \leq d$, the set defined in (4), $\mathcal{D}(g_1,g_2,\ldots,g_r)$, is an open set in $\mathbb{R}^d_+ / \{0\}$. Define $f(x) = ||x_{g_i}|| + \sum_{i \neq j} |x_{g_i} - \bar{x}_{g_j}|$, which is a continuous and convex function. It is differentiable at $x$ if and only if $x \in \mathcal{D}(g_1,g_2,\ldots,g_r)$. Therefore, by Lemma 4, $f \circ \lambda$ is convex, and it is differentiable at $A$ if and only if $\lambda(A) \in \mathcal{D}(g_1,g_2,\ldots,g_r)$. Since $\lambda \in \mathcal{M}(g_1,g_2,\ldots,g_r)$ is the pre-image of the open set $\mathbb{R}^d_+ / \{0\}$ under a continuous function $h \circ \lambda$, where

$$h(x) = \prod_{i \neq j} |x_{g_i} - \bar{x}_{g_j}||x_{g_i}|.$$  

Therefore, it is open.
Appendix B.6  Proof of Theorem 1

Proof. Throughout the proof, we adopt the usual localization procedure as detailed in Jacod and Protter (2012). Note that

\[ V(\Delta_n, X; F) = k_n \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} f \left( \tilde{\lambda}_{ik_n \Delta_n} \right) = k_n \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} (f \circ \lambda)(\tilde{c}_{ik_n \Delta_n}). \]

By Assumption 2 and Lemma 4, \( f \circ \lambda \) is a continuous vector-valued function. Moreover, for \( c \in \mathcal{M}_d^+ \), \( \| f \circ \lambda(c) \| \leq K (1 + \| \lambda(c) \|^5) \leq K (1 + \| c \|^5) \). Below we prove this theorem for any spectral function \( F \) that is bounded by \( K (1 + \| c \|^5) \).

We start with a function \( F \) bounded by \( K \) everywhere. We extend the definition of \( \tilde{c} \) to the entire interval \([0, t]\) by letting:

\[ \tilde{c}_s = \tilde{c}_{(i-1)k_n \Delta_n}, \quad \text{for} \quad (i-1)k_n \Delta_n \leq s < ik_n \Delta_n. \]

Note that for any \( t > 0 \), we have

\[
E \left\| V(\Delta_n, X; F) - \int_0^t F(c_s) ds \right\| \\
\leq k_n \Delta_n \left( [t/(k_n \Delta_n)] + \int_0^{[t/(k_n \Delta_n)]} \| F(\tilde{c}_s) - F(c_s) \| ds + \int_{[t/(k_n \Delta_n)]}^t \| F(c_s) \| ds \right) \\
\leq K k_n \Delta_n + \int_0^{[t/(k_n \Delta_n)]} \| F(\tilde{c}_s) - F(c_s) \| ds.
\]

By the fact that \( \tilde{c}_s - c_s \xrightarrow{p} 0 \), it follows that \( E \| F(\tilde{c}_s) - F(c_s) \| \to 0 \), which is bounded uniformly in \( s \) and \( n \) because \( F \) is bounded. Therefore, by the dominated convergence theorem, we obtain the desired convergence.

Next we show the convergence holds under the polynomial bound on \( F \). Denote \( \psi \) to be a \( C^\infty \) function on \( \mathbb{R}^+ \) such that \( 1_{[1, \infty)}(x) \leq \psi(x) \leq 1_{[1/2, \infty)}(x) \). Let \( \psi_\varepsilon(c) = \psi(\| c \|/\varepsilon) \), and \( \psi_\varepsilon'(c) = 1 - \psi_\varepsilon(c) \). Since the function \( F \cdot \psi_\varepsilon'(c) \) is continuous and bounded, the above argument implies that \( V(\Delta_n, X; F \cdot \psi_\varepsilon'(c)) \xrightarrow{p} \int_0^t F \cdot \psi_\varepsilon'(c_s) ds \), for any fixed \( \varepsilon \). When \( \varepsilon \) is large enough, we have \( \int_0^t F \cdot \psi_\varepsilon'(c_s) ds = \int_0^t F(c_s) ds \) by localization, since \( c_s \) is locally bounded. On the other hand, \( F \cdot \psi_\varepsilon(c) \leq K \| c \|^{\xi} 1_{\{\| c \| \geq \varepsilon \}} \), for \( \varepsilon > 1 \). So it remains to show that

\[
\lim_{\varepsilon \to 0} \lim_{n \to \infty} \E \left( k_n \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} \| \tilde{c}_{ik_n \Delta_n} \|^{\xi} 1_{\{\| \tilde{c}_{ik_n \Delta_n} \| > \varepsilon \}} \right) = 0.
\]

By (9.4.7) of Jacod and Protter (2012), there exists some sequence \( a_n \) going to 0, such that

\[
\E \left( \| \tilde{c}_{ik_n \Delta_n} \|^{\xi} 1_{\{\| \tilde{c}_{ik_n \Delta_n} \| > \varepsilon \}} \right) \leq \frac{K}{\varepsilon^\xi} + Ka_n \Delta_n^{(1-\xi + \omega(2\xi-\gamma))},
\]

which establishes the desired result.
Appendix B.7 Proof of Proposition 1

Proof. We divide the proof into several steps. To start, we need some additional notations. Let $X'$ and $c'$ denote the continuous parts of the processes $X$ and $c$, respectively. Also, we introduce $\hat{c}'_{ikn}\Delta_n$ to denote the estimator constructed similarly as in (5) with $X$ replaced by $X'$ and without truncation, namely

$$\hat{c}'_{ikn}\Delta_n = \frac{1}{k\Delta_n} \sum_{j=1}^{k_n} (\Delta_{ikn+j}X')^T(\Delta_{ikn+j}X').$$

In addition, $\hat{\lambda}'_{ikn}\Delta_n$ corresponds to the vector of eigenvalues of $\hat{c}'_{ikn}\Delta_n$, and

$$V'(\Delta_n, X; F) = \Delta_n \sum_{i=0}^{t/(k\Delta_n)} f\left(\hat{\lambda}'_{ikn}\Delta_n\right).$$

We also define

$$\check{c}_{ikn}\Delta_n = \frac{1}{k\Delta_n} \int_{ikn\Delta_n}^{(i+1)k\Delta_n} c_s ds, \quad \check{\beta}_{ikn}^n = \hat{c}'_{ikn}\Delta_n - c_{ikn}\Delta_n, \quad \alpha^n = (\Delta^n X')(\Delta^n X')^T - c\Delta_n\Delta_n, \quad \text{and}$$

$$\eta^n_i = \left( E \left( \sup_{i\Delta_n \leq u \leq (i+1)\Delta_n} \left| b_i\Delta_n + u - b_i\Delta_n \right|^2 F_i \Delta_n \right) \right)^{1/2}.$$

We first collect some known estimates in the next lemma:

**Lemma 6.** Under the assumptions of Proposition 1, we have

$$E \left( \sup_{0 \leq u \leq s} \| c_{t+u} - c_t \|^q | F_t \right) \leq K s^{1+q/2}, \quad \| E (c_{t+s} - c_t | F_t) \| \leq K s, \quad \text{(B.1)}$$

$$E \| (\Delta^n X)(\Delta^n X)^T \left\{ \Delta^n X \left| \leq u_n \right. \right\} - (\Delta^n X')(\Delta^n X')^T \| \leq K a_n \Delta_n^{(2-\gamma)\omega+1}, \text{ for some } a_n \to 0, \quad \text{(B.2)}$$

$$E \left( \| \hat{c}_{ikn}\Delta_n - \hat{c}'_{ikn}\Delta_n \| \right)^q \leq K a_n \Delta_n^{(2q-\gamma)\omega+q-1}, \text{ for some } q \geq 1, \text{ and } a_n \to 0, \quad \text{(B.3)}$$

$$E \left( \| \hat{c}'_{ikn}\Delta_n - \hat{c}_{ikn}\Delta_n \| \right)^p \leq K k^{-p/2}, \text{ for some } p \geq 1, \quad \text{(B.4)}$$

$$E \left( \| \alpha^n_i \| | F_i \Delta_n \right) \leq K \Delta_n^q, \text{ for some } q \geq 0, \quad \text{(B.5)}$$

$$\| E (\alpha^n_i | F_i \Delta_n) \| \leq K \Delta_n^{3/2} \left( \Delta_n^{1/2} + \eta_i^n \right)^2, \quad \text{(B.6)}$$

$$\left| E \left( \alpha^n_i c_{ikn}^n \right) \right| - \left( \left( c_{ikn}^n + c_{ikn}^n \right) \Delta^n_i \right) \leq K \Delta_n^{5/2}, \quad \text{(B.7)}$$

$$\| E \left( \beta^n_{ikn} | F_{ikn} \Delta_n \right) \| \leq K \Delta_n \frac{1}{\Delta_n} \left( \Delta_n^{1/2} + \eta_i^n \right), \quad \text{(B.8)}$$

$$\| E \left( \beta^n_{ikn} | F_{ikn} \Delta_n \right) \| \leq K \Delta_n \frac{1}{\Delta_n} \left( \Delta_n^{1/2} + \eta_i^n \right), \quad \text{(B.9)}$$

$$E \left( \| \beta^n_{ikn} \| | F_{ikn} \Delta_n \right) \leq K \left( k^{-q/2} + k \Delta_n \right), \text{ for some } q \geq 2, \quad \text{(B.10)}$$

$$\Delta_n \sum_{i=1}^{t/(k\Delta_n)} \eta_i^n \rightarrow 0. \quad \text{(B.11)}$$
Proof of Lemma 6. These estimates are given by Lemma A.2 in Li and Xiu (2016), (4.3), (4.8), (4.10), (4.11), (4.12), (4.18), Lemmas 4.2 and 4.3 of Jacod and Rosenbaum (2013), and Lemma 13.2.6 of Jacod and Protter (2012).

Now we return to the proof of Proposition 1.

1) We show that we can restrict the domain of function $f$ to some compact set, where both the estimates $\{\bar{c}_{ikn}\}_{i=0,1,2,...,[t/(k_n\Delta_n)]}$ and the sample path of $\{c_s\}_{s \in [0,t]}$ take values. By (B.4), we have for $p \geq 1$,

$$\mathbb{E}\left[\sup_{0 \leq i \leq [t/(k_n\Delta_n)]} \|\bar{c}'_{ikn\Delta_n} - \bar{c}_{ikn\Delta_n}\|^p\right] \leq Kk_n^{-p/2}.$$  

Therefore, by the maximal inequality, we deduce, by picking $p > 2/\gamma - 2$,

$$\mathbb{E}\left|\sup_{0 \leq i \leq [t/(k_n\Delta_n)]} \|\bar{c}'_{ikn\Delta_n} - \bar{c}_{ikn\Delta_n}\|^p\right| \leq K\Delta_n^{-1}k_n^{-p/2-1} \rightarrow 0,$n

therefore, $\sup_{0 \leq i \leq [t/(k_n\Delta_n)]} \|\bar{c}'_{ikn\Delta_n} - \bar{c}_{ikn\Delta_n}\| = o_p(1)$. Moreover, by (B.2) we have

$$\mathbb{E}\left(\sup_{0 \leq i \leq [t/(k_n\Delta_n)]} \|\bar{c}'_{ikn\Delta_n} - \bar{c}_{ikn\Delta_n}\|\right) \leq \frac{1}{\Delta_n} \sum_{i=1}^{[t/\Delta_n]-k_n} \mathbb{E}\left(\|\bar{c}_{ikn\Delta_n} - \bar{c}_{ikn\Delta_n}\|\right) \leq K\Delta_n^{(2-\gamma)\omega - 1+\epsilon} \rightarrow 0.$$

As a result, we have as $\Delta_n \rightarrow 0$,

$$\sup_{0 \leq i \leq [t/(k_n\Delta_n)]} \|\bar{c}_{ikn\Delta_n} - \bar{c}_{ikn\Delta_n}\| \rightarrow 0.$$  

Note that by Assumption 3, for $0 \leq s \leq t$, $c_s \in C \cap \mathcal{M}^\omega(g_1, g_2, \ldots, g_r)$, where $C$ is a convex and open set. Therefore, $\{\bar{c}_{ikn\Delta_n}\}_{i=0,1,2,...,[t/(k_n\Delta_n)]} \in C$ by convexity. For $n$ large enough, $\{\bar{c}_{ikn\Delta_n}\}_{i=0,1,2,...,[t/(k_n\Delta_n)]} \in C$, with probability approaching 1, by (B.12). Since $C \subset \mathcal{M}(g_1, g_2, \ldots, g_r)$, we can restrict the domain of $f$ to the compact set $\lambda(\bar{C}) \subset \mathcal{D}(g_1, g_2, \ldots, g_r)$, in which $f$ is $C^\infty$ with bounded derivatives. Moreover, because $\lambda_{g_j}(\cdot), 1 \leq j \leq r$ are continuous functions, $\min_{1 \leq j \leq r-1}(\lambda_{g_j}(\cdot) - \lambda_{g_{j+1}}(\cdot))$ is hence continuous, so that $\inf_{c \in \mathcal{C}}\{\min_{1 \leq j \leq r-1}(\lambda_{g_j}(c) - \lambda_{g_{j+1}}(c))\} \geq \epsilon > 0$. It follows from Lemma 4 and Theorem 3.5 of Silhavý (2000) that $F(\cdot)$ is $C^\infty$ with bounded derivatives on $C$.

2) Next, we have

$$\|V(\Delta_n, X; F) - V'(\Delta_n, X; F)\| \leq k_n\Delta_n \sum_{i=0}^{[t/(k_n\Delta_n)]} \|F(\bar{c}_{ikn\Delta_n}) - F(\bar{c}'_{ikn\Delta_n})\| \leq Kk_n\Delta_n \sum_{i=0}^{[t/(k_n\Delta_n)]} \|\bar{c}_{ikn\Delta_n} - \bar{c}'_{ikn\Delta_n}\|.$$  

By (B.3), we have

$$\mathbb{E}\left(\|\bar{c}_{ikn\Delta_n} - \bar{c}'_{ikn\Delta_n}\|\right) \leq K\Delta_n^{(2-\gamma)\omega}.$$  

51
where $a_n$ is some sequence going to 0, as $n \to \infty$, which implies

$$V(\Delta_n, X; F) - V(\Delta_n, X; F) = O_p(a_n \Delta_n^{(2-\gamma)\infty}).$$  \hfill (B.13)

As a result, given the conditions on $\infty$, we have

$$k_n \left( V(\Delta_n, X; F) - V(\Delta_n, X; F) \right) = o_p(1),$$

hence we can proceed with $V'$ in the sequel.

3) Then we show for each $1 \leq h \leq d$, we have

$$k_n \left( V'(\Delta_n, X; F_h) - k_n \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} \left( F_h(c_{ik_n} \Delta_n) - \frac{1}{2k_n} \sum_{j,k,l,m=1}^{d} \partial_{jk,lm}^2 F_h(c_{ik_n} \Delta_n) (c_{jl,ik_n} \Delta_n c_{lm,ik_n} \Delta_n + c_{jm,ik_n} \Delta_n c_{kl,ik_n} \Delta_n) \right) \right) = o_p(1).$$

where $F_h$ is the $h$th entry of the vector-valued function $F$.

To prove it, we decompose the left hand side into 4 terms:

$$R^n_{1,h} = k_n^{2 \Delta_n} \sum_{i=0}^{[t/(k_n \Delta_n)]} \left( F_h(c_{ik_n} \Delta_n + \beta_{ik_n}^{(n)}) - F_h(c_{ik_n} \Delta_n) - \sum_{l,m=1}^{d} \partial_{lm} F_h(c_{ik_n} \Delta_n) \beta_{ik_n}^{(n,lm)} \right) \right) \right) \right), \hfill (B.14)

$$R^n_{2,h} = k_n^{2 \Delta_n} \sum_{i=0}^{[t/(k_n \Delta_n)]} \frac{1}{2} \sum_{j,k,l,m=1}^{d} \partial_{jk,lm}^2 F_h(c_{ik_n} \Delta_n) \left( \beta_{ik_n}^{(n,lm)} \beta_{ik_n}^{(n,jk)} - \frac{1}{k_n} (c_{jl,ik_n} \Delta_n c_{lm,ik_n} \Delta_n + c_{jm,ik_n} \Delta_n c_{kl,ik_n} \Delta_n) \right), \hfill (B.15)

$$R^n_{3,h} = k_n^{2 \Delta_n} \sum_{i=0}^{[t/(k_n \Delta_n)]} \sum_{l,m=1}^{d} \partial_{lm} F_h(c_{ik_n} \Delta_n) \sum_{u=1}^{k_n} (c_{lm,(ik_n+u)} \Delta_n - c_{lm,ik_n} \Delta_n) \hfill (B.16)

$$R^n_{4,h} = k_n \sum_{i=0}^{[t/(k_n \Delta_n)]} \sum_{l,m=1}^{d} \partial_{lm} F_h(c_{ik_n} \Delta_n) \sum_{u=1}^{k_n} \beta_{ik_n+u}^{(n,lm)} \hfill (B.17)

We first consider $R^n_{1,h}$. By (B.10), we have

$$E(|R^n_{1,h}|) \leq K k_n^{2 \Delta_n} \sum_{i=0}^{[t/(k_n \Delta_n)]} E \| \beta_{ik_n}^{(n)} \|^3 \leq K (k_n^{2 \Delta_n} \sum_{i=0}^{[t/(k_n \Delta_n)]} (k_n^{-3/2} + k_n \Delta_n) \hfill \leq K k_n^{2 \Delta_n} + K k_n^{-1/2} \to 0.$$
As to \( R^n_{2,h} \), we denote the term inside the summation of \( R^n_{2,h} \) as \( \nu^n_{ikn} \). So we have

\[
R^n_{2,h} = k_n^2 \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} \left( \nu^n_{ikn} - E(\nu^n_{ikn} | \mathcal{F}_{ikn} \Delta_n) + E(\nu^n_{ikn} | \mathcal{F}_{ikn} \Delta_n) \right).
\]

By (B.9) we have

\[
|E(\nu^n_{ikn} | \mathcal{F}_{ikn} \Delta_n)| \leq K \Delta_n^{1/2} \left( k_n^{-1/2} + k_n \Delta_n^{1/2} + \eta^n_{ikn} \right).
\]

On the other hand, by (B.10), we can derive

\[
E \left( \nu^n_{ikn} - E(\nu^n_{ikn} | \mathcal{F}_{ikn} \Delta_n) \right)^2 \leq 3 k_n \Delta_n.
\]

Then Doob’s inequality implies that

\[
E \left( \sup_{s \leq t} \left| \sum_{i=0}^{[s/(k_n \Delta_n)]} \left( \nu^n_{ikn} - E(\nu^n_{ikn} | \mathcal{F}_{ikn} \Delta_n) \right) \right| \right) \leq K t.
\]

As a result, by (B.11)

\[
E \left( |R^n_{2,h}| \right) \leq k_n^2 \Delta_n \sup_{s \leq t} \left| \sum_{i=0}^{[s/(k_n \Delta_n)]} \left( \nu^n_{ikn} - E(\nu^n_{ikn} | \mathcal{F}_{ikn} \Delta_n) \right) \right| + k_n^2 \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} |E(\nu^n_{ikn} | \mathcal{F}_{ikn} \Delta_n)|
\]

\[
\leq K k_n^2 \Delta_n + K k_n \sqrt{\Delta_n} \left( k_n \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} \eta_{ikn} \right) \rightarrow 0.
\]

The proof for \( E(|R^n_{3,h}|) \rightarrow 0 \) is similar. Denote the term inside the summand as \( \xi^n_{ikn} \). By (B.1) and the Cauchy-Schwarz inequality, we have

\[
|E(\xi^n_{ikn} | \mathcal{F}_{ikn} \Delta_n)| \leq K k_n^2 \Delta_n, \quad E \left( |\xi^n_{ikn}|^2 | \mathcal{F}_{ikn} \Delta_n \right) \leq K k_n^3 \Delta_n.
\]

By Doob’s inequality again,

\[
E \left( |R^n_{3,h}| \right) \leq k_n \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} E \left( |E(\xi^n_{ikn} | \mathcal{F}_{ikn} \Delta_n)| \right) + k_n \Delta_n \left( \sum_{i=0}^{[t/(k_n \Delta_n)]} E \left( |\xi^n_{ikn}|^2 \right) \right)^{1/2}
\]

\[
\leq K k_n^2 \Delta_n \rightarrow 0.
\]

For \( R^n_{4,h} \), it can be shown in the proof of Theorem 2 below that \( R^n_{4,h} = O_p(k_n \sqrt{\Delta_n}) = o_p(1) \).

4) Finally, it is sufficient to show that

\[
k_n \left( \sum_{i=0}^{[t/(k_n \Delta_n)]} \left( \int_{ikn \Delta_n}^{(i+1)k_n \Delta_n} F_h(c_{ikn} \Delta_n) ds - \int_{ikn \Delta_n}^{(i+1)k_n \Delta_n} F_h(c_s) ds \right) - \int_{ikn \Delta_n}^{t} \left( \int_{ikn \Delta_n}^{[t/(k_n \Delta_n)]k_n \Delta_n} F_h(c_s) ds \right) \right) \overset{p}{\rightarrow} 0,
\]

as the similar result holds if we replace \( F_h(c_{ikn} \Delta_n) \) by
\[
\partial_{jk,lm}^2 F_h(c_{ij} \Delta_n) (c_{jl,ik} \Delta_n c_{km,ik} \Delta_n + c_{jm,ik} \Delta_n c_{kl,ik} \Delta_n) \]. Since \( F_h \) is bounded, the second term is bounded by \( Kk_n^2 \Delta_n \to 0 \). As to the first term, we notice that
\[
\zeta_n^j \kappa_k = \int_{i\kappa_n \Delta_n}^{(i+1)\kappa_n \Delta_n} F_h(c_{ik} \Delta_n) \, ds - \int_{i\kappa_n \Delta_n}^{(i+1)\kappa_n \Delta_n} F_h(c_s) \, ds
\]
is measurable with respect to \( \mathcal{F}_{(i+1)\kappa_n \Delta_n} \), and that
\[
|E(\zeta_n^j | \mathcal{F}_{i\kappa_n \Delta_n})| \leq K(k_n \Delta_n)^2, \quad E(|\zeta_n^j|^2 | \mathcal{F}_{i\kappa_n \Delta_n}) \leq K(k_n \Delta_n)^3,
\]
so the same steps as in (2) and (3) yield the desired results. \( \square \)

**Appendix B.8 Proof of Theorem 2**

**Proof.** To start, we decompose
\[
\frac{1}{\sqrt{\Delta_n}} \left( \tilde{V}(\Delta_n, X; F) - k_n \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} F(c_{i \kappa_n \Delta_n}) \right) = \frac{1}{k_n \sqrt{\Delta_n}} (R_1^n + R_2^n + R_3^n + R_4^n + R_6^n),
\]
where \( R_i^n = \left( R_{i,1}^n, R_{i,2}^n, \ldots, R_{i,d}^n \right)^T \), for \( i = 1, 2, 3, 4, \) and \( 5 \), with \( R_{1,h}^n, R_{2,h}^n, R_{3,h}^n \) and \( R_{4,h}^n \) given by equations (B.14) - (B.17). In addition, \( R_{5,h}^n \) and \( R_{6,h}^n \) are given by
\[
R_{5,h}^n = \frac{k_n \Delta_n}{2} \sum_{i=0}^{[t/(k_n \Delta_n)]} \sum_{j,k,l,m=1}^d \left( \partial_{jk,lm}^2 F_h(c_{i \kappa_n \Delta_n}) (c_{jl,ik} \Delta_n c_{km,ik} \Delta_n + c_{jm,ik} \Delta_n c_{kl,ik} \Delta_n) \right.
\]
\[
- \partial_{jk,lm}^2 F_h(c_{i \kappa_n \Delta_n}) (c_{jl,ik} \Delta_n c_{km,ik} \Delta_n + c_{jm,ik} \Delta_n c_{kl,ik} \Delta_n) \left.) \right).
\]
\[
R_{6,h}^n = \frac{k_n \Delta_n}{2} \sum_{i=0}^{[t/(k_n \Delta_n)]} \sum_{j,k,l,m=1}^d \left( \partial_{jk,lm}^2 F_h(c_{i \kappa_n \Delta_n}) (c_{jl,ik} \Delta_n c_{km,ik} \Delta_n + c_{jm,ik} \Delta_n c_{kl,ik} \Delta_n) \right.
\]
\[
- \partial_{jk,lm}^2 F_h(c_{i \kappa_n \Delta_n}) (c_{jl,ik} \Delta_n c_{km,ik} \Delta_n + c_{jm,ik} \Delta_n c_{kl,ik} \Delta_n) \left.) \right) + k_n \left( V(\Delta_n, X; F) - V'(_i\Delta_n, X; F) \right).
\]
We have shown in the proof of Proposition 1 that \( R_{i}^n = o_p(k_n \sqrt{\Delta_n}) \), for \( i = 1, 2, 3 \). Therefore, these terms do not contribute to the asymptotic variance of \( \tilde{V}(\Delta_n, X; F) \).

Next, we show that \( R_{5,h}^n \) is also \( o_p(k_n \sqrt{\Delta_n}) \). By (B.10) and the mean-value theorem, we have
\[
E|R_{5,h}^n| \leq K k_n \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} E \left| c_{i \kappa_n \Delta_n} - c_{i \kappa_n \Delta_n}' \right| \leq K (k_n^{-1/2} + k_n \Delta_n) = o_p(k_n \sqrt{\Delta_n}).
\]

As to \( R_{6,h}^n \), by (B.13) and the mean-value theorem, we have
\[
E|R_{6,h}^n| \leq K k_n \Delta_n \sum_{i=0}^{[t/(k_n \Delta_n)]} E \left| \tilde{c}_{i \kappa_n \Delta_n} - \tilde{c}_{i \kappa_n \Delta_n}' \right| = O_p(a_n \Delta_n^{2-\gamma}) = o_p(k_n \sqrt{\Delta_n}).
\]

Hence, \( \omega \geq \frac{1-\varepsilon}{2-\gamma} > \frac{1}{4-2\gamma} \) is sufficient to warrant the desired rate.
As a result, except for the term that is related to $R_{4n}$, all the remainder terms on the right-hand side of (B.18) vanish. We write $R_{4n}$ as
\[ R_{4n} = k_n \sum_{i=1}^{d} \omega_{i,n}^{n,i} \alpha_{i,n}^{n,i}, \] where $\omega_{i,n}^{n,i}$ is a vector measurable with respect to $\mathcal{F}_{(i-1)\Delta_n}$, and $\|\omega_{i,n}^{n,i}\| \leq K$. To prove the stable convergence result, we start with
\[ \frac{1}{\sqrt{\Delta_n}} \sum_{i=0}^{[t/(k_n\Delta_n)]k_n} \omega_{i,n}^{n,i} E\left(\alpha_{i,n}^{n,i} | \mathcal{F}_{i\Delta_n}\right) \leq \frac{1}{\sqrt{\Delta_n}} \sum_{i=0}^{[t/(k_n\Delta_n)]k_n} K\Delta_n^{3/2}(\sqrt{\Delta_n} + E(\eta_i^{n,i})) \to 0, \] where we use (B.6) and (B.11). Moreover, by (B.5), we have
\[ \frac{1}{\Delta_n^2} \sum_{i=0}^{[t/(k_n\Delta_n)]k_n} ||\omega_{i,n}^{n,i}||^4 E\left(\|\alpha_{i,n}^{n,i}\|^4 | \mathcal{F}_{i\Delta_n}\right) \leq K\Delta_n \to 0. \] Also, similar to (4.18) in Jacod and Rosenbaum (2013), we have $E\left(\alpha_{i,n}^{n,i}\Delta_i^{n} N | \mathcal{F}_{i\Delta_n}\right) = 0$, for $N = W$ or any $N$ that is an arbitrary bounded martingale orthogonal to $W$, which readily implies
\[ \frac{1}{\Delta_n} \sum_{i=0}^{[t/(k_n\Delta_n)]k_n} \omega_{i,n}^{n,i} E\left(\alpha_{i,n}^{n,i}\Delta_i^{n} N | \mathcal{F}_{i\Delta_n}\right) \stackrel{p}{\to} 0. \] Finally, note that for any $1 \leq p, q \leq d$, by (B.7),
\[ \frac{1}{\Delta_n} \sum_{i=0}^{[t/(k_n\Delta_n)]k_n} |\omega_{i,p}^{n,jk} \omega_{i,q}^{n,lm}| E\left(\alpha_{i,n}^{n,jk} \alpha_{i,n}^{n,lm} | \mathcal{F}_{i\Delta_n}\right) - (c_{i\Delta_n,j,l} c_{i\Delta_n,k,m} + c_{i\Delta_n,j,m} c_{i\Delta_n,k,l}) \Delta_n^2 \leq K\Delta_n^{1/2}, \] which implies
\[ \frac{1}{\Delta_n} \sum_{i=0}^{[t/(k_n\Delta_n)]k_n} \omega_{i,p}^{n,jk} \omega_{i,q}^{n,lm} E\left(\alpha_{i,n}^{n,jk} \alpha_{i,n}^{n,lm} | \mathcal{F}_{i\Delta_n}\right) = \frac{1}{\Delta_n^2} \sum_{i=0}^{[t/(k_n\Delta_n)]k_n} \omega_{i,p}^{n,jk} \omega_{i,q}^{n,lm} \left(c_{i\Delta_n,j,l} c_{i\Delta_n,k,m} + c_{i\Delta_n,j,m} c_{i\Delta_n,k,l}\right) \Delta_n^2 + o_p(1) \stackrel{p}{\to} \int_0^t \partial_{jk} F_p(c_s) \partial_{lm} F_q(c_s) \left(c_{s,j,l} c_{s,k,m} + c_{s,j,m} c_{s,k,l}\right). \] Finally, by Theorem IX.7.28 of Jacod and Shiryaev (2003), we establish
\[ \frac{1}{k_n\sqrt{\Delta_n}} R_{4n} \overset{L^1}{\rightarrow} \mathcal{W}_t, \]
where \( \mathcal{W}_t \) is conditional Gaussian on an extension of the probably space, with a covariance matrix

\[
E (\mathcal{W}_{p,t} | \mathcal{W}_{q,t} | \mathcal{F}) = \sum_{j,k,l,m=1}^d \int_0^t \partial_{jk} F_p(c_s) \partial_{lm} F_q(c_s) \left( c_{s,ji} c_{s,km} + c_{s,jm} c_{s,kl} \right). 
\]

**Appendix B.9  Proof of Proposition 2**

*Proof.* As we have seen from the above proof, we have for any \( c \in \mathcal{C} \),

\[
\| \partial_{jk} F_p(c) \partial_{lm} F_q(c) (c_{jl} c_{km} + c_{jm} c_{kl}) \| \leq K (1 + \|c\|^2),
\]

which, combined with the same argument in the proof of Theorem 1, establishes the desired result.

**Appendix B.10  Proof of Corollary 1**

*Proof.* The first statement on consistency follows immediately from Theorem 1, as Assumption 2 holds with \( \zeta = 1 \). Next, we prove the central limit result. For any \( 1 \leq p \leq d \), we define \( f^\lambda_p \) as,

\[
f^\lambda_p(\bar{x}) = \frac{1}{g_p - g_{p-1}} \sum_{j=g_{p-1}+1}^{g_p} x_j,
\]

and \( f^\lambda \) is \( C^\infty \) and Lipchitz. By Lemma 4 we can derive

\[
\partial_{jk} F^\lambda_p(c_s) = \sum_{u=1}^d O_{uj} \partial_{u,\lambda} F^\lambda_p(c_s) O_{uk} = \frac{1}{g_p - g_{p-1}} \sum_{u=1}^d \sum_{v=g_{p-1}+1}^{g_p} O_{uj} e^v \partial_{u,\lambda} F^\lambda_p(\bar{x}), \quad \partial^2 f^\lambda_p(\bar{x}) = 0,
\]

Therefore, the asymptotic covariance matrix is given by

\[
\int_0^t \sum_{j,k,l,m=1}^d \partial_{jk} F^\lambda_p(c_s) \partial_{lm} F^\lambda_q(c_s) (c_{jl} c_{km} + c_{jm} c_{kl}) \, ds \\
= \frac{1}{g_p - g_{p-1}} \frac{1}{g_q - g_{q-1}} \int_0^t \sum_{j,k,l,m=1}^d \sum_{u=g_{p-1}+1}^{g_p} \sum_{v=g_{q-1}+1}^{g_q} O_{uj} O_{vk} O_{ul} O_{um} (c_{jl} c_{km} + c_{jm} c_{kl}) \, ds \\
= \frac{2}{(g_p - g_{p-1})(g_q - g_{q-1})} \int_0^t \sum_{l,m=1}^d \sum_{u=g_{p-1}+1}^{g_p} \sum_{v=g_{q-1}+1}^{g_q} O_{el} O_{vm} O_{ul} O_{um} \lambda^2_{v,s} \, ds \\
= \frac{2}{(g_p - g_{p-1})(g_q - g_{q-1})} \int_0^t \sum_{v=g_{p-1}+1}^{g_p} \sum_{u=g_{q-1}+1}^{g_q} \delta_{u,v} \lambda^2_{v,s} \, ds
\]

56
\[= \frac{2\delta_{p,q}}{(g_p - g_{p-1})} \int_0^t \chi_{g_p,s}^2 ds. \quad (B.19)\]

Next, we calculate the bias-correction term. Recall that the estimator is given by

\[F_p^\lambda (\hat{c}_{ik\alpha}) = \frac{1}{g_p - g_{p-1}} \sum_{v=g_{p-1}+1}^{g_p} \hat{\lambda}_{v,ik\alpha},\]

where \(\hat{\lambda}_{v,ik\alpha}\) is the corresponding eigenvalue of the sample covariance matrix \(\hat{c}_{ik\alpha}\). Although \(\hat{c}_{ik\alpha}\) and \(c_{ik\alpha}\) may have different eigenstructure, it is easy to verify that the functional forms of the second order derivative of \(F_p^\lambda\) evaluated at both points turn out to be the same, so here we only provide the calculations based on \(\hat{c}_{ik\alpha}\). Since almost surely, sample eigenvalues are simple, it implies from Lemma 4 that

\[\partial_{jk,lm}^2 F_p^\lambda (\hat{c}_{ik\alpha}) = \sum_{u,v=1}^d A^F_{uv}(\lambda(\hat{c}_{ik\alpha})) \hat{O}_{ul} \hat{O}_{uj} \hat{O}_{vk} \hat{O}_{vm}\]

\[= \frac{1}{g_p - g_{p-1}} \sum_{h=g_{p-1}+1}^{g_p} \sum_{u,v=1, u \neq v}^d \frac{e^h_u - e^h_v}{\lambda_{u,ik\alpha} - \lambda_{v,ik\alpha}} \hat{O}_{ul} \hat{O}_{uj} \hat{O}_{vk} \hat{O}_{vm}\]

\[= \frac{1}{g_p - g_{p-1}} \sum_{h=g_{p-1}+1}^{g_p} \sum_{u=1, u \neq h}^d \frac{1}{\lambda_{h,ik\alpha} - \hat{\lambda}_{u,ik\alpha}} \left( \hat{O}_{ul} \hat{O}_{uj} \hat{O}_{hk} \hat{O}_{hm} + \hat{O}_{uh} \hat{O}_{hj} \hat{O}_{uk} \hat{O}_{um} \right),\]

where \(\hat{O}\) is the orthogonal matrix such that \(\hat{O}_{ik\alpha} \hat{O}^\top = \text{Diag}(\lambda(\hat{c}_{ik\alpha}))\). The dependence of \(\hat{O}\) on \(ik\alpha\) is omitted for brevity.

To facilitate the implementation, we consider the matrix \(\hat{\lambda}_{h,ik\alpha} - \hat{\lambda}_{1,ik\alpha} \cdot \hat{\lambda}_{h,ik\alpha} - \hat{\lambda}_{2,ik\alpha} \cdot \hat{\lambda}_{h,ik\alpha} - \hat{\lambda}_{d,ik\alpha}\). Note that

\[\text{Diag}(\hat{\lambda}_{h,ik\alpha} - \hat{\lambda}_{1,ik\alpha}, \hat{\lambda}_{h,ik\alpha} - \hat{\lambda}_{2,ik\alpha}, \cdots, \hat{\lambda}_{h,ik\alpha} - \hat{\lambda}_{d,ik\alpha}) = \hat{O}(\hat{\lambda}_{h,ik\alpha} - \hat{c}_{ik\alpha} \hat{\lambda}_{h,ik\alpha} \hat{O}^\top),\]

hence we have

\[\left(\hat{\lambda}_{h,ik\alpha} - \hat{c}_{ik\alpha}\right)^+ = \hat{O}^\top \text{Diag} \left( \frac{1}{\hat{\lambda}_{h,ik\alpha} - \hat{\lambda}_{1,ik\alpha}}, \frac{1}{\hat{\lambda}_{h,ik\alpha} - \hat{\lambda}_{2,ik\alpha}}, \cdots, 0, \cdots, \frac{1}{\hat{\lambda}_{h,ik\alpha} - \hat{\lambda}_{d,ik\alpha}} \right) \hat{O}.\]

As a result, we obtain

\[\left(\hat{\lambda}_{h,ik\alpha} - \hat{c}_{ik\alpha}\right)_+ = \sum_{u=1, u \neq p}^d \frac{1}{\hat{\lambda}_{h,ik\alpha} - \hat{\lambda}_{u,ik\alpha}} \hat{O}_{ul} \hat{O}_{um},\]

Therefore, we have

\[\partial_{jk,lm}^2 F_p^\lambda (\hat{c}_{ik\alpha}) = \frac{1}{g_p - g_{p-1}} \sum_{h=g_{p-1}+1}^{g_p} \hat{O}_{hl} \left(\hat{\lambda}_{h,ik\alpha} - \hat{c}_{ik\alpha}\right)_+ \hat{O}_{hm} \hat{O}_{hj} \left(\hat{\lambda}_{h,ik\alpha} - \hat{c}_{ik\alpha}\right)_+ \hat{O}_{dl}.\]
Now we can calculate the following term, which is used for bias-correction:

\[
\sum_{j,k,l,m=1}^{d} \partial_{jklm}^p F_{p}^{\lambda} (\widehat{c}_{i \Delta_n}) (\widehat{c}_{j \Delta_n} \widehat{c}_{km,i \Delta_n} + \widehat{c}_{jm,i \Delta_n} \widehat{c}_{kl,i \Delta_n})
\]

\[
= \frac{1}{g_p - g_{p-1}} \sum_{h=g_{p-1}+1}^{g_p} \sum_{j,k,l,m=1}^{d} \left( \partial_{hk} \left( \widehat{\lambda}_{h,i \Delta_n} - \widehat{\gamma}_{i \Delta_n} \right)_{jl} + \partial_{hm} \left( \widehat{\lambda}_{h,i \Delta_n} - \widehat{\gamma}_{i \Delta_n} \right)_{kl} \right)
\]

\[
\cdot (\widehat{c}_{j \Delta_n} \widehat{c}_{km,i \Delta_n} + \widehat{c}_{jm,i \Delta_n} \widehat{c}_{kl,i \Delta_n})
\]

\[
= \frac{2}{g_p - g_{p-1}} \sum_{h=g_{p-1}+1}^{g_p} \widehat{\lambda}_{h,i \Delta_n} \text{Tr} \left( \left( \widehat{\lambda}_{h,i \Delta_n} - \widehat{\gamma}_{i \Delta_n} \right) \widehat{\gamma}_{i \Delta_n} \right).
\]

(B.20)

The last equality uses the following observation:

\[
\widehat{\lambda}_{h,i \Delta_n} \widehat{\gamma}_{i \Delta_n} + \widehat{\gamma}_{i \Delta_n} \widehat{\gamma}_{i \Delta_n} = 0,
\]

which concludes the proof of (ii). The proof of (iii) uses the same calculations as above. Note that we can apply Theorem 2 with only Assumption 4, because the spectral function here only depends on \( \lambda_g \).

**Appendix B.11  Proof of Proposition 3**

**Proof.** By Lemma 5 and the uniform convergence of \( \widehat{c}_{ik \Delta_n} - \widehat{\gamma}_{ik \Delta_n} \rightarrow 0 \) established above, we can restrict the domain of \( \gamma_g(\cdot) \) to the set \( C \), in which it is \( C^\infty \) with bounded derivatives. By Theorem 21 of Protter (2004), we have

\[
\frac{[t/(k_n \Delta_n)]-1}{t} \sum_{i=1}^{[t/(k_n \Delta_n)]-1} \gamma^T_{g,ik_n \Delta_n} (X_{i+1}k_n \Delta_n - X_{ik_n \Delta_n}) \xrightarrow{u.c.p} \int_0^t \gamma_{g,s} dX_s.
\]

Therefore, it remains to show that

\[
\frac{[t/(k_n \Delta_n)]-1}{t} \sum_{i=1}^{[t/(k_n \Delta_n)]-1} \left( \widehat{\gamma}^T_{g,(i-1)k_n \Delta_n} - \gamma^T_{g,ik_n \Delta_n} \right) (X_{i+1}k_n \Delta_n - X_{ik_n \Delta_n}) \xrightarrow{u.c.p} 0.
\]

Define a \( \mathcal{F}_{(i+1)k_n \Delta_n} \)-measurable function:

\[
\xi_{ik_n} = \left( \widehat{\gamma}^T_{g,(i-1)k_n \Delta_n} - \gamma^T_{g,ik_n \Delta_n} \right) (X_{i+1}k_n \Delta_n - X_{ik_n \Delta_n}).
\]

By standard estimates in (B.1) with \( c \) replaced by \( X \), (B.3), and (B.10),

\[
E(\xi_{ik_n} | \mathcal{F}_{ik_n \Delta_n}) = E(\widehat{\gamma}_{g,(i-1)k_n \Delta_n} - \gamma_{g,ik_n \Delta_n} |_{\mathcal{F}_{ik_n \Delta_n}}) \leq KE(\widehat{c}_{(i-1)k_n \Delta_n} - \gamma_{ik_n \Delta_n} |_{(k_n \Delta_n)})
\]

\[
\leq K \left( (k_n \Delta_n)^{1/2} + a_n \Delta_n^{(2-\gamma)/2} + \sqrt{k_n^{-1} + k_n \Delta_n} \right) (k_n \Delta_n)
\]

See page 160 of Magnus and Neudecker (1999).
Moreover, we have by the same estimates above,
\[
E\left( |\xi_{ik_n}|^2 | \mathcal{F}_{ik_n,\Delta_n} \right) \leq (k_n \Delta_n + a_n \Delta_n^{(d-\gamma)\over 2} + k_n^{-1} + k_n \Delta_n)k_n \Delta_n.
\]

Finally, using Doob’s inequality, and measurability of \( \xi_{ik_n} \), we obtain
\[
E \left( \sup_{0 \leq s \leq t} \left| \sum_{i=1}^{[t/(k_n \Delta_n)]-1} \xi_{ik_n} \right| \right)
\leq \sum_{i=1}^{[t/(k_n \Delta_n)]-1} E|\xi_{ik_n}| \mathcal{F}_{ik_n \Delta_n}
\leq \sum_{i=1}^{[t/(k_n \Delta_n)]-1} E|\xi_{ik_n}| \mathcal{F}_{ik_n \Delta_n}
\leq K \left( (k_n \Delta_n)^{1/2} + a_n \Delta_n^{(2-\gamma)\over 2} + \sqrt{k_n^{-1} + k_n \Delta_n} \right) + K(k_n \Delta_n + a_n \Delta_n^{(d-\gamma)\over 2} + k_n^{-1} + k_n \Delta_n)^{1/2}
\to 0,
\]
because \((d-\gamma)\over 2 \geq 1\) under our assumptions on \( \overline{\gamma} \) and \( \overline{\zeta} \), which establishes the proof. \( \square \)

**Appendix B.12 Proof of Corollary 2**

*Proof*. The \((p, q)\) entry of the asymptotic covariance matrix is given by
\[
\int_0^t \sum_{j,k,l,m=1}^d \partial_{jk} \gamma_{gp,s} \partial_{lm} \gamma_{gs,s} (c_{jl,s} c_{km,s} + c_{jm,s} c_{kl,s}) \, ds
= \int_0^t \sum_{j,k,l,m=1}^d (\lambda_{g,s} (c_{jl,s} c_{km,s} + c_{jm,s} c_{kl,s}) \, ds
= \int_0^t \sum_{j=1}^d (\lambda_{g,s} \gamma_{gs,s} + \gamma_{gs,s} \gamma_{gs,s} c_{jm,s} c_{kl,s}) \, ds
= \int_0^t \lambda_{g,s} (\lambda_{g,s} \gamma_{gs,s} + \gamma_{gs,s} c_{jm,s} c_{kl,s}) \, ds,
\]
where we use \((\lambda_{g,s} \gamma_{gs,s} + \gamma_{gs,s} c_{km,s} + c_{jm,s} c_{kl,s}) = 0\), and \(\sum_{k=1}^d \gamma_{gk,s} c_{km,s} = \lambda_{g,s} \gamma_{gs,s} \). To calculate the asymptotic bias, we note that the \( \widehat{\gamma}_{s} \) has only simple eigenvalues almost surely. Denote \( \tilde{\lambda}_{h} \) and \( \tilde{\gamma}_{h} \) as the corresponding eigenvalue and eigenvector. We omit the dependence on time \( s \) to simplify the notations. By Lemma 2, we obtain
\[
\sum_{j,k,l,m=1}^d \partial_{jk,lm}^2 \widehat{\gamma}_{g,s} (\widehat{c}_{jl} \widehat{c}_{km} + \widehat{c}_{jm} \widehat{c}_{kl})
= -2 \sum_{p \neq g} \frac{\widehat{\lambda}_{p} \widehat{\gamma}_{p}}{(\lambda_{g} - \lambda_{p})^2} \sum_{l,m=1}^d \widehat{\gamma}_{p,l} \widehat{\gamma}_{g,m} + \sum_{p \neq g} \frac{\widehat{\lambda}_{p} \widehat{\gamma}_{g}}{(\lambda_{g} - \lambda_{p})^2} \left( \sum_{l,m=1}^d \widehat{\gamma}_{p,l} \widehat{\gamma}_{g,m} \widehat{\gamma}_{p,m} + \sum_{l,m=1}^d \widehat{\gamma}_{g,l} \widehat{\gamma}_{p,m} \widehat{\gamma}_{p,l} \right)
\]
59
Since \(\gamma_g(\cdot)\) is a \(C^\infty\) function, it is straightforward using the proof of Theorem 2 that the desired CLT holds, even though \(\gamma_g(\cdot)\) is not a spectral function.

\[\sum \sum_{p \neq g \neq p} \frac{\lambda_p \lambda_q}{(\lambda_g - \lambda_p)(\lambda_g - \lambda_q)} \left( \sum_{l,m=1}^{d} \hat{\gamma}_{ql} \hat{\gamma}_{pm} \hat{\gamma}_{gm} \hat{\gamma}_{qh} + \sum_{l,m=1}^{d} \hat{\gamma}_{pm} \hat{\gamma}_{ql} \hat{\gamma}_{ql} \hat{\gamma}_{qh} \right) + \sum \sum_{p \neq g \neq p} \frac{\lambda_q \lambda_g}{(\lambda_g - \lambda_p)(\lambda_g - \lambda_q)} \left( \sum_{l,m=1}^{d} \hat{\gamma}_{ql} \hat{\gamma}_{gm} \hat{\gamma}_{ph} \hat{\gamma}_{pm} + \sum_{l,m=1}^{d} \hat{\gamma}_{pm} \hat{\gamma}_{ql} \hat{\gamma}_{ql} \hat{\gamma}_{ph} \right) + \sum \sum_{p \neq g \neq p} \frac{\lambda_p \lambda_q}{(\lambda_g - \lambda_p)(\lambda_g - \lambda_q)} \left( \sum_{l,m=1}^{d} \hat{\gamma}_{pm} \hat{\gamma}_{ql} \hat{\gamma}_{gm} \hat{\gamma}_{ph} + \sum_{l,m=1}^{d} \hat{\gamma}_{pm} \hat{\gamma}_{ql} \hat{\gamma}_{ql} \hat{\gamma}_{ph} \right) = - \sum \frac{\lambda_p \lambda_g}{(\lambda_g - \lambda_p)^2} \hat{\gamma}_{qh}.\]

\textbf{Appendix B.13 Proof of Propositions 4 and 5}

Proof. This follows by applying the “delta” method, using Lemma 2 and Theorem 13.2.4 in Jacod and Protter (2012).