Functional Time Series Approach to Analysing Asset Returns Co-movements

Patrick Wongsa-art

Cardiff Business School
Cardiff University
United Kingdom

Yingcun Xia

Department of Statistics and Applied Probability
National University of Singapore
Singapore

Abstract:
We introduce a new approach for modeling the time varying behavior and time series evolution of asset returns co-movements. Here, the co-movement in each period is captured by a trajectory of returns correlation, then a sequence of this over time and the time series evolution are studied. We rely on functional principal components to achieve dimension reduction and to construct the dynamic space of interest, while introducing a new class of information criteria in order to identify the finite dimensionality of the curve time series. Our method is able to combine two of the most applied ideas in the literature, namely economics (or finance) based and time-series based time-varying correlation models. This offers a general specification that is able to model processes of time-varying time-series correlations generated under many existing models that have dominated the financial literature for several decades. To illustrate its empirical relevance, we apply our method to model the time varying co-movement of exchange rate returns for a group of small open economies with large financial sectors. Our empirical results indicate that concepts of time varying correlation enabled by existing methods are too restrictive to accommodate fully the time varying behavior and time series evolution of the returns correlation. On the other hand, our method gives a more complete picture and is able to provide more accurate correlation forecasts.

1. Introduction

In the disciplines of economics and finance, co-movements between returns on financial assets are believed to carry many important implications. For instance, information about co-movements among international stock returns are needed to determine gains from international portfolio diversification when optimizing a portfolio. Also, a calculation of minimum variance hedge ratio needs updated information on the co-movements between returns of assets in the hedge.

It is also well-known that such co-movements are time-varying. Overall, there are generally two main approaches to explaining the time-varying behavior. On the one hand, many studies follow the Engle's (2002) Dynamic Conditional Correlation (DCC) idea, which imposes the GARCH-type dynamics on returns correlation (see e.g. Cappiello et al. (2006) and Kasch and Caporin (2013)), and generalize it to obtain models that allow asset return co-movements to be directly explained as a deterministic function of time (e.g. Aslanidis and Casas (2013)). On the other hand, a number of studies put forward market variables, such as measures of return and/or volatility, as keys determinants of returns correlation (see e.g. Ang and Chen (2002), Hafner et al. (2006), Silvernoinen and Terävirta (2015) and Jiang et al. (2016)).

In this paper, we introduce a new approach that can take these two ideas into consideration simultaneously. To the best of our knowledge, Kasch and Caporin (2013) is the only work that attempts to combine economics (or finance) based and time-series based time-varying correlation models. They introduce the Threshold Generalized DCC (T-GDCC) model by directly introducing a threshold structure in either the DCC-GARCH specification or its asymmetric generalized DCC (GDCC) extension (Cappiello and Engle (2006)) to allow for the effects of returns volatility. Our approach differs significantly from these existing models.

We take the view that co-movements between a pair of asset returns can be explained entirely by a trajectory of the returns' correlation. The time-series evolution and serial dependence of such trajectories are captured by a functional process that is constructed in Section 2.1 as a combination of a time-invariant and a time-varying components. Here, the former is analogous to the concept of returns co-movement assumed in the Semparametric Correlation (SP-C) model of Hafner et al. (2006). Whereas, the latter is constructed by a stochastic process, which summaries the dynamics of the functional process in question. In this regard, we assume that the time-varying component admits the Karhunen-Loéve expansion, which is the stochastic parallel of the Fourier Expansion. However, unlike in traditional functional data analysis, which focuses on the covariance function as the Mercer kernel, this paper explores the use of the auto-covariance function. Analogously to the well-known Portmanteau test procedure in time series analysis, we focus our analysis on the first p lags, where p is a small integer. We consider an alternative linear operator, which can be intuitively viewed as the summation of the auto-covarainces, in order to empirically construct the Karhunen-Loéve expansion. The resulting procedure is not only able to address previous limitations of functional data analysis (FDA) in financial applications (see, for example, Müller et al. (2011)), but also offers a general specification that can model processes of time-varying correlations generated under many existing models, which have dominated the financial literature for several decades.

The first obstacle that we must face is the fact that the above-mentioned trajectory of returns' correlation is not observable in practice. To address this, we treat the correlation coefficient of asset returns for each period (e.g. for each day), as a correlation trajectory that is assumed to be a realisation of the functional time series of interest. In Section 2.2, we introduce the local-linear estimator for such a correlation coefficient for each day by making use of the within-day returns (e.g. 1-minute or 5-minute returns). Accordingly, we establish an estimator for the linear operator, which was discussed in the previous paragraph. Performing eigenanalysis in the Hilbert space is not a trivial matter, however. In Section 2.3, we discuss an alternative method, which transforms the problem into an eigenanalysis for a finite matrix. Such method is based on suggestions made in Bathia et al. (2010) and Benko et al. (2008).

Section 2.4 focuses on asymptotic results. Firstly, we presents the uniform convergence rate for the local linear estimator mentioned in the above point. The uniform convergence is essential in our study since it ensures that the estimated functional correlation is close to the true function everywhere. We also present asymptotic results for the proposed estimation procedure. The proof of these deviates quite significantly from existing studies in functional data analysis. The key to such a difference is the interaction between nonparametrics and the operator theory used in this work. In addition, these results hold for a process with an infinite order of the Karhunen-Loéve expansion.

The key to the practicality of our method is its ability to construct the dynamic space for the functional correlation time series of interest. Our approach relies on functional principal components. When principal component analysis is involved, dimension reduction is achieved naturally and the truncated Karhunen-Loéve expansion becomes our main focus. In this paper, we present a set of theoretical results that help to verify the use of the truncated expansion as an acceptable approximation. Firstly, we establish the consistency of such a representation by showing that if the dimension is allowed to increase to infinity, then the mean squared error using the finite representation in the space of the deterministic function converges to zero. Secondly, we establish its optimality by showing that among all truncated expansions of the same form, the truncated Karhunen-Loéve expansion minimises the integrated mean squared error. Moreover, we introduce in Section 3 a new class of information criteria to help to identify the finite dimensionality of the curve time series. We present the consistency of our selection and show that it also holds for the case in which the dimensionality tends to infinity.

To illustrate its empirical relevance, we conduct a series of simulation studies in Section 4 and apply our analytical framework to model time varying correlation of exchange rate returns for a group of small open economies with large financial sectors, namely the United Kingdom, Switzerland, Norway and Sweden, in Section 5. Here, let us summarize some important findings. Our empirical results indicate that concepts of time varying correlation enabled by existing methods, e.g. the SP-C and the DCC-GARCH models, might still be too rigid to accommodate fully the time-varying behavior and temporal evolution of the returns correlation. The SP-C model, for example, does not allow functional variation of the correlation over time and is therefore not able to provide accurate in-sample forecasts of the functional correlation when compared to our method. In addition, the GARCH-type evolution offered by models within the DCC-GARCH family may not be able to capture the time series evolution of the correlation that truly takes place. Our empirical results suggests that the time series evolution of returns correlation involves both low frequency cycles with relatively lengthy periodicity and trend, and high frequency cycles (say, for example, the day-of-the-week effects) with a shorter periodicity.

Finally, Section 6 concludes. All the technical discussion and proofs are relegated to the Appendix.

2. Functional Correlation Time Series

Throughout this paper, let t and τ denote two different indexes. For instance, in the empirical analysis presented in Section 5 we assume that within the t^{th} day there are discrete grid of time points

$$t_\tau = \tau\Delta, \quad \tau = 1, \dots, m$$

in which $m = \lfloor I/\Delta \rfloor$, where I denotes the overall length of time interval and $\lfloor Q \rfloor$ stands for the largest integer smaller than or equal to Q. In this regard, the motivating daily trading data are recorded on a regular grid often with Δ quantified as either 5 or 1 minute, such that $\Delta \to 0$ signifies higher frequency trading data and implies that $m \to \infty$.

Moreover, by letting $P_{k,t,\tau}$ be the price of asset k at the τ^{th} time point in the t^{th} day, then $r_{k,t,\tau} = p_{k,t,\tau} - p_{k,t,\tau-1}$ is the log-return, i.e. the continuous compounded return, by which $p_{k,t,\tau} = \ln(P_{k,t,\tau})$. If they are relevant, the log-return of other assets, such as ℓ, can also be similarly defined. In the analysis that follows, we assume that returns follow

$$r_{k,t,\tau} = \mu_{k,t}(U_{t,\tau}) + \sigma_{k,t}(U_{t,\tau})\epsilon_{k,t,\tau} \quad \text{and} \quad r_{\ell,t,\tau} = \mu_{\ell,t}(U_{t,\tau}) + \sigma_{\ell,t}(U_{t,\tau})\epsilon_{\ell,t,\tau},$$

where $E\{\epsilon_{k,t,\tau}|U_{t,\tau}\} = E\{\epsilon_{\ell,t,\tau}|U_{t,\tau}\} = 0$ and $E\{\epsilon^2_{j,t,\tau}|U_{t,\tau}\} = 1$ almost surely. Clearly, $r_{k,t,\tau}$ and $r_{\ell,t,\tau}$ depend on $U_{t,\tau}$, but this dependence is omitted from the notation to simplify exposition. Assumption 7.1 in the appendix discuss the probability and time series properties of $r_{k,t,\tau}$, $r_{\ell,t,\tau}$ and $U_{t,\tau}$ in detail.

The correlation coefficient formulated in (2.1) below portrays the concept of co-movement that we are interested in, i.e. the correlation between a pair of returns as driven by $U_{t,\tau}$,

$$Corr_t\{r_{k,t,\tau}, r_{\ell,t,\tau}|U_{t,\tau} = u\} = \frac{\mu_{k\ell,t}(u) - \mu_{\ell,t}(u)\mu_{k,t}(u)}{\sqrt{\sigma^2_{\ell,t}(u)\sigma^2_{k,t}(u)}} \tag{2.1}$$

for $u \in \mathcal{I}$, where $\mu_{k\ell,t}(u) = E\{r_{k,t,\tau} r_{\ell,t,\tau} | U_{t,\tau} = u\}$, $\sigma_{k,t}(u)$ is positive over u in the support of $U_{t,\tau}$ and $\mathcal{I}$ signifies a compact interval. Since this is simply $E[\epsilon_{k,t,\tau}\epsilon_{\ell,t,\tau} | U_{t,\tau} = u]$, we are indeed modeling the time series evolution of the error covariance, where $U_{t,\tau}$ can be any financial or economic variables.

When a given day t is considered, providing availability of the returns in high frequency trading (e.g. based on closing prices that are recorded every 1 min), we should be able to formulate consistent estimates of $Corr_t\{r_{k,t,\tau}, r_{\ell,t,\tau} | U_{t,\tau} = u\}$ for all $t = 1, \ldots, n$. However, these estimates are not capable of explaining the time-varying behavior of the correlation. In this paper, we are interested in the time series evolution of the trajectory that explains the returns correlation with respect to $U_{t,\tau}$. To this end we propose expressing the correlation process as a combination of a time-invariant and stochastic time-varying components. This idea is congruent with well-known existing models (e.g. the DCC-GARCH, GDCC and the T-GDCC) and will be thoroughly discussed in the next section.

2.1. Basic Construction

Let $\rho_1(u), \ldots, \rho_n(u)$ denote the functional time series defined on a compact interval $\mathcal{I}$. In this paper, we take the view that such functional process expresses the time series evolution of the the trajectory of returns correlation. Moreover,

$$\rho_t(u) = \varrho(u) + \vartheta_t(u), \quad u \in \mathcal{I}, \tag{2.2}$$

where $\mathcal{I}$ signifies a compact interval, $\varrho(u) = E\{\rho_t(u)\}$ takes into account the possible non-time-varying part and $\vartheta_t(u)$ is the stochastic process that drives the time-varying component. In addition, we assume that $\rho_t(u)$ takes values in $\mathcal{L}_2(\mathcal{I})$, i.e. the Hilbert space consisting of all square integrable functions defined on $\mathcal{I}$ with the inner product

$$\langle f, g \rangle = \int_{\mathcal{I}} f(u) g(u) du, \quad f, g \in \mathcal{L}_2(\mathcal{I}). \tag{2.3}$$

In this regard, $\rho_t(u)$ depicts the instantaneous correlation of the returns, whereas $\rho_1(u), \ldots, \rho_n(u)$ form a strictly stationary time series process hereafter referred to as "functional correlation time series" (FC-TS). Assumption 7.2 in Appendix 7.4 discusses the strict stationarity and mixing properties in detail. It follows from the definition of stationarity, that a stationary time series should fluctuate around a constant level. Hence, for the stationary FC-TS, the level $\varrho(u) = E\{\rho_t(u)\}$ can be seen as the equilibrium value, while deviations from the mean $\vartheta_t(u)$ can be interpreted as deviations from equilibrium.

A similar concept of time-variation was also studied by Müller et al. (2011), but within the context of the volatility (see also Dalla et al. (2015) for a similar treatment on the mean). Here, the FC-TS expresses a concept of time-varying correlations, while also providing a convenient vehicle to accommodate such a nonstationary feature into a stationary setup. In addition, such a formulation of the correlation is more general that it can handle processes of time-varying correlations of time series suggested by existing models, which have dominated financial literature for many years. For example, the popular DCC-GARCH specification is resulted when return correlations are independent of $U_{t,\tau}$ and evolve temporally under the GARCH-type time series evolution. The correlation curve $\rho_t(u)$ is reduced simply to a step function under the T-GDCC specification. By rewriting (2.2) as perturbation $\rho_t(u_{t,\tau}) - \varrho(u_{t,\tau}) = \vartheta_t(u_{t,\tau})$, the concept of correlation considered by Hafner et al. (2006) is obtained when the time-varying component (i.e. the right hand side) is zero. We shall revisit these points and present some empirical illustration in Section 5.

We now explain the construction of the functional process in (2.2) in detail. We begin with a common approach in functional data analysis; particularly by assuming that the continuous covariance function

$$M^{(0)}(u,v) = Cov\{\rho_t(u), \rho_t(v)\}, \tag{2.4}$$

defined on $\mathcal{I} \times \mathcal{I}$, is the Mercer kernel satisfying the Fredholm integral equation

$$\int_{\mathcal{I}} M^{(0)}(u,v)\varphi_j(v)dv = \lambda_j \varphi_j(u), \quad j \geq 1, \tag{2.5}$$

where λ_j and $\varphi_j(u)$ respectively are eigenvalues and orthogonal eigenfunctions (i.e. $\langle \varphi_i, \varphi_j \rangle = 1$ for $i = j$, and 0 otherwise) of the compact symmetric linear operator $\mathrm{M}^{(0)}$ on $\mathcal{L}_2(\mathcal{I})$. In this respect, $\vartheta_t(u)$ is a zero-mean square-integrable stochastic process indexed over $\mathcal{I}$ also with the continuous covariance function $M^{(0)}(u,v)$. Under these conditions, the Karhunen-Loéve Theorem suggests that we may decompose

$$\vartheta_t(u) = \sum_{j=1}^{\infty} \xi_{tj}\varphi_j(u), \qquad \xi_{tj} = \int_{\mathcal{I}} \vartheta_t(u)\varphi_j(u)du, \tag{2.6}$$

where $E(\xi_{tj}) = 0$, $Var(\xi_{tj}) = \lambda_j$ and $Cov(\xi_{ts}, \xi_{tj}) = 0$ for $s \neq j$ (see e.g. Yao et al. (2005a,b), Hall and Vial (2006), Wang (2008) and the references therein). Furthermore, $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, in the other words; the only possible limit point of a sequence of eigenvalues is 0.

The decomposition in (2.6) carries various important methodological and empirical implications. We focus here on the former and revisit the latter point in Section 3. On the one hand, $M^{(0)}(u,v)$ can be expressed as

$$M^{(0)}(u,v) = \sum_{j=1}^{\infty} \lambda_j \varphi_j(u)\varphi_j(v) \tag{2.7}$$

by venture of the Mercer's theorem. In their study of daily functional volatility, Müller et al. (2011) associated to $M^{(0)}(u,v)$ the linear operator $\mathrm{M}^{(0)}$ and solved the Fredholm integral equation (2.5). Nonetheless, doing so assumes that the process in question is temporally uncorrelated. To account for this, the authors must make an empirical compromise by randomly selecting only a sub-sample of days in order to enhance the temporal independence.

In this paper, we shall take a different approach. Let

$$M^{(q)}(u,v) \equiv Cov\{\rho_t(u), \rho_{t+q}(v)\} \tag{2.8}$$

denote the continuous auto-covariance function defined on $\mathcal{I} \times \mathcal{I}$ for any $q \neq 0$. Analogously to (2.7), we can formulate based on (2.6) the auto-covariance function

$$\begin{aligned} M^{(q)}(u,v) &= E\left\{\left(\sum_{i=1}^{\infty} \xi_{ti}\varphi_i(u)\right)\left(\sum_{j=1}^{\infty} \xi_{t+q,j}\varphi_j(v)\right)\right\} \\ &= \sum_{i,j=1}^{\infty} \sigma_{ij}^{(q)} \varphi_i(u)\varphi_j(v) \end{aligned} \tag{2.9}$$

defined on $\mathcal{I} \times \mathcal{I}$, in which

$$\sigma_{ij}^{(q)} = E\{\xi_{ti}\xi_{t+q,j}\}$$

denotes the autocovariance at lag q for $i = j$ and cross-autocovariance for $i \neq j$. For any $f \in \mathcal{L}_2(\mathcal{I})$ and $\mathrm{M}^{(\mathrm{q})} f \in \mathcal{L}_2(\mathcal{I})$, let

$$(\mathrm{M}^{(\mathrm{q})} f)(u) = \int_{\mathcal{I}} M^{(q)}(u,v) f(v) dv \tag{2.10}$$

such that the linear operator $\mathrm{M}^{(\mathrm{q})}$ is compact and may be decomposed as

$$\mathrm{M}^{(\mathrm{q})} = \sum_{i,j=1}^{\infty} \sigma_{ij}^{(q)} \varphi_i \otimes \varphi_j. \tag{2.11}$$

Or equivalently,

$$(\mathrm{M}^{(\mathrm{q})} f)(u) = \sum_{i,j=1}^{\infty} \sigma_{ij}^{(q)} \langle \varphi_j, f\rangle \varphi_i(u). \tag{2.12}$$

These, together with (2.6), suggest that by focusing on $M^{(q)}(u,v)$ and $\mathrm{M}^{(\mathrm{q})}$ (instead of $M^{(0)}(u,v)$ and $\mathrm{M}^{(0)}$) the dynamics (i.e. the time series evolution) of the FC-TS can be explained entirely by that of the vector process $\boldsymbol{\xi}_t = (\xi_{t1}, \xi_{t2} \ldots)'$.

Analogously to the well-known Portmanteau test procedure in the time series analysis, we suggest focusing on

$$M(u,v) = \sum_{1 \le q \le p} M^{(q)}(u,v), \tag{2.13}$$

where p is a pre-specified positive integer. Under the strict stationarity and mixing properties outlined in Appendix 7.4, p can be specified as a small positive integer in practice since the serial dependence should decay quickly as the lag increases. However, this idea is ineffective since it may not necessarily be the case that

$$\int_{\mathcal{I}} \sum_{1 \le q \le p} M^{(q)}(u,v) f(v)\ dv \neq 0. \tag{2.14}$$

This is due to the fact that $\mathrm{M}^{(\mathrm{q})}$ is not a nonnegative definite operator unlike $\mathrm{M}^{(0)}$. In other words, ones cannot ensure that

$$\langle \mathrm{M}^{(\mathrm{q})} f, f \rangle \;=\; \sum_{i,j=1}^{\infty} \sigma_{ij}^{(q)} \int_{\mathcal{I}} \left(\int_{\mathcal{I}} \varphi_j(v) f(v)\ dv \right) \varphi_i(u) f(u)\ du \tag{2.15}$$

is greater than or equal 0 since $\sigma_{ij}^{(q)}$ are the autocovariances at lag q.

To address this problem, we follow the suggestion made by Bathia et al. (2010) and employ an alternative operator K whereby

$$K(u,v) = \textstyle\sum_{q=1}^{p} N^{(q)}(u,v) \tag{2.16}$$

$$N^{(q)}(u,v) = \textstyle\int_{\mathcal{I}} M^{(q)}(u,z) M^{(q)}(v,z) dz = \sum_{i,j=1}^{\infty} w_{ij}^{(q)} \varphi_i(u) \varphi_j(v) \tag{2.17}$$

and $\boldsymbol{W}^{(q)} = \left(w_{ij}^{(q)} \right) = \boldsymbol{\Sigma}^{(q)} \boldsymbol{\Sigma}^{(q)\prime}$ is a nonnegative definite matrix. In this regard,

$$\begin{aligned} (\mathrm{N}^{(\mathrm{q})} f)(u) \;&=\; \int N^{(q)}(u,v) f(v)\ dv \\ &=\; \sum_{i,j=1}^{\infty} w_{ij}^{(q)} \langle \varphi_i, f \rangle \varphi_j(u) \;=\; (\mathrm{M}^{(\mathrm{q})} \mathrm{M}^{(\mathrm{q})*} f)(u), \end{aligned} \tag{2.18}$$

where $\mathrm{M}^{(\mathrm{q})*}$ signifies the adjoint of $\mathrm{M}^{(\mathrm{q})}$. This suggests that $\mathrm{N}^{(\mathrm{q})} = \mathrm{M}^{(\mathrm{q})} \mathrm{M}^{(\mathrm{q})*}$ and also that

$$\mathrm{Im}(\mathrm{N}^{(\mathrm{q})}) = \mathrm{Im}(\mathrm{M}^{(\mathrm{q})} \mathrm{M}^{(\mathrm{q})*}),$$

where Im signifies the image of the operator (see Appendix 7.1 for detailed definitions). In addition, K is a nonnegative definite operator since

$$\begin{aligned} \langle \mathrm{N}^{(\mathrm{q})} f, f \rangle \;&=\; \sum_{i,j=1}^{\infty} w_{ij}^{(q)} \left(\int_{\mathcal{I}} \varphi_i(u) f(u)\ du \right) \left(\int \varphi_j(v) f(v)\ dv \right) \\ &=\; \langle \mathrm{M}^{(\mathrm{q})*} f, \mathrm{M}^{(\mathrm{q})*} f \rangle \end{aligned} \tag{2.19}$$

where $(\mathrm{M}^{(\mathrm{q})*} f)(u) = \int_{\mathcal{I}} M^{(q)}(v,u) f(v)\ dv$. Furthermore:

Lemma 2.1. *Let $\{\psi_j(u)\}_{j=1}^{\infty}$ denote the orthonormal eigenfunctions of* K *and θ_j signify the corresponding eigenvalue to the eigenfunction $\psi_j(u)$. The relation* $\mathrm{K}\psi_j = \theta_j \psi_j$ *holds and*

$$\mathcal{V}_t(u) = \lim_{d \to \infty} \sum_{j=1}^{d} \eta_{tj} \psi_j(u) \ \textit{uniformly}, \tag{2.20}$$

where $\eta_{tj} = \int_{\mathcal{I}} \mathcal{V}_t(u)\psi_j(u)\ du$, *in the sense that*

$$E(\mathcal{V}_t(u) - \sum_{j=1}^{d} \eta_{tj}\psi_j(u))^2 \to 0. \tag{2.21}$$

While the proof of Lemma 2.1 is presented in Appendix 7.2, the validity of using $\mathcal{V}_t(u)$ instead of $\vartheta_t(u)$ will be made clear in Section 3.

2.2. Estimators

This section and the next focus on estimation aspects of the concepts introduced in the previous section. Firstly, by following a common practice in functional data analysis, we may define the estimator of $M^{(q)}(u,v)$ as

$$\tilde{M}^{(q)}(u,v) = \frac{1}{n-p}\sum_{j=1}^{n-p}\{\rho_j(u) - \tilde{\varrho}(u)\}\{\rho_{j+q}(v) - \tilde{\varrho}(v)\}, \tag{2.22}$$

where $\tilde{\varrho}(u) = n^{-1}\sum_{1\le j\le n}\ \rho_j(u)$ is the estimator of the expected correlation. Accordingly, the estimator for $K(u,v)$ can be written as

$$\tilde{K}(u,v) = \sum_{q=1}^{p}\tilde{N}^{(q)}(u,v) = \sum_{q=1}^{p}\int_{\mathcal{I}}\tilde{M}^{(q)}(u,z)\tilde{M}^{(q)}(v,z)\ dz. \tag{2.23}$$

However, these require observing the FC-TS, which is usually not possible in practice. To address this issue, we propose using $Corr_t\{r_{k,t,\tau}, r_{\ell,t,\tau}|U_{t,\tau} = u\}$ to represent a trajectory that is assumed to be a realization of the stochastic function $\rho_t(u)$.

To this end, we rely on the formula in (2.1) to construct the needed estimator. In particular, our nonparametric estimator of the correlation is constructed as

$$\hat{\rho}_t(u) = \frac{\hat{\mu}_{k\ell,t}(u) - \hat{\mu}_{\ell,t}(u)\hat{\mu}_{k,t}(u)}{\sqrt{\hat{\sigma}^2_{\ell,t}(u)\hat{\sigma}^2_{k,t}(u)}}, \tag{2.24}$$

where $\hat{\mu}_{k\ell,t}(u)$, $\hat{\mu}_{k,t}(u)$, $\hat{\mu}_{\ell,t}(u)$, $\hat{\sigma}^2_{k,t}(u)$ and $\hat{\sigma}^2_{\ell,t}(u)$ denote local-linear estimators of $\mu_{k\ell,t}(u)$, $\mu_{k,t}(u)$, $\mu_{\ell,t}(u)$, $\sigma^2_{k,t}(u)$ and $\sigma^2_{\ell,t}(u)$, respectively. In a general sense, these local-linear estimators are obtained based on the following minimisation problem

$$\arg\min_{\beta_0,\beta_1}\sum_{\tau=1}^{m}\{y_{t,\tau} - \beta_0 - \beta_1(U_{t,\tau} - u)\}^2\,\kappa_h(U_{t,\tau} - u),$$

where $\kappa_h(U_{t,\tau} - u) = \kappa\big(\frac{U_{t,\tau}-u}{h}\big)/h$, $\kappa(\cdot)$ is a kernel function and h is the bandwidth parameter. $y_{t,\tau}$ is either $r_{k,t,\tau}r_{\ell,t,\tau}$, $r_{k,t,\tau}$, $r_{\ell,t,\tau}$, $(r_{k,t,\tau} - \hat{\mu}_{k,t}(u))^2$ or $(r_{\ell,t,\tau} - \hat{\mu}_{\ell,t}(u))^2$. By letting

$$W_{t,\tau}(u) = \frac{W_{m,h}(U_{t,\tau} - u)}{\sum_{\tau=1}^{m}W_{m,h}(U_{t,\tau} - u)}, \tag{2.25}$$

where $W_{m,h}(U_{t,\tau} - u) = s_{m,h,2}\kappa_h(U_{t,\tau} - u) - s_{m,h,1}\kappa_h(U_{t,\tau} - u)(U_{t,\tau} - u)$ and $s_{m,h,r} = \sum_{\tau=1}^{m} \kappa_h(U_{t,\tau} - u)(U_{t,\tau} - u)^r$ (for $r = 0, 1, 2$), these local-linear estimators can be formulated as follows $\hat{\mu}_{k\ell,t}(u) = W_{t,\tau}(u) r_{k,t,\tau} r_{\ell,t,\tau}$, $\hat{\mu}_{k,t}(u) = W_{t,\tau}(u) r_{k,t,\tau}$, $\hat{\mu}_{\ell,t}(u) = W_{t,\tau}(u) r_{\ell,t,\tau}$, $\hat{\sigma}^2_{k,t}(u) = W_{t,\tau}(u)(r_{k,t,\tau} - \hat{\mu}_{k,t}(u))^2$ and $\hat{\sigma}^2_{\ell,t}(u) = W_{t,\tau}(u)(r_{\ell,t,\tau} - \hat{\mu}_{\ell,t}(u))^2$.

Moreover, by replacing the time series $\rho_1(u), \ldots, \rho_n(u)$ with $\hat{\rho}_1(u), \ldots, \hat{\rho}_n(u)$, the estimators $\tilde{M}^{(q)}(u,v)$ and $\tilde{K}(u,v)$ can be respectively replaced by

$$\hat{M}^{(q)}(u,v) = \frac{1}{n-q} \sum_{j=1}^{n-q} \{\hat{\rho}_j(u) - \hat{\varrho}(u)\}\{\hat{\rho}_{j+q}(v) - \hat{\varrho}(v)\}, \tag{2.26}$$

where

$$\hat{\varrho}(u) = \frac{1}{n} \sum_{1 \le j \le n} \hat{\rho}_j(u), \tag{2.27}$$

and

$$\begin{aligned} \hat{K}(u,v) &= \sum_{q=1}^{p} \int_{\mathcal{I}} \hat{M}^{(q)}(u,z) \hat{M}^{(q)}(v,z) dz \qquad (2.28) \\ &= \frac{1}{(n-p)^2} \sum_{t,s=1}^{n-p} \sum_{q=1}^{p} \{\hat{\rho}_t(u) - \hat{\varrho}(u)\}\{\hat{\rho}_s(v) - \hat{\varrho}(v)\}\langle \hat{\rho}_{t+q} - \hat{\varrho}, \hat{\rho}_{s+q} - \hat{\varrho}\rangle. \end{aligned}$$

2.3. Eigenanalysis

Performing eigenanalysis in the Hilbert space is not a trivial matter. To this end, Bathia et al. (2010) suggest transforming the problem into an eigenanalysis for a finite matrix by making use of the well-known duality method introduced in Benko et al. (2008). To follow the Bathia et al. (2010) approach, we begin with the infeasible, i.e *"tilde"*, version as done in the previous section.

Let us view the curves $\rho_t(u) - \tilde{\varrho}(u)$ and $\rho_{t+q}(u) - \tilde{\varrho}(u)$ as $\infty \times 1$ vectors denoted by $\tilde{\boldsymbol{\varrho}}_t$ and $\tilde{\boldsymbol{\varrho}}_{t+q}$, respectively. Also, let $\tilde{\boldsymbol{\varrho}}_t' \tilde{\boldsymbol{\varrho}}_{t+q} = \langle \rho_t - \tilde{\varrho}, \rho_{t+q} - \tilde{\varrho}\rangle$, $\tilde{\mathcal{Y}}_q = (\tilde{\boldsymbol{\varrho}}_{1+q}, \ldots, \tilde{\boldsymbol{\varrho}}_{n-p+q})$ and $\tilde{\mathcal{Y}}_q' = (\tilde{\boldsymbol{\varrho}}_{1+q}, \ldots, \tilde{\boldsymbol{\varrho}}_{n-p+q})'$. Then, $\tilde{K}(u,v)$ can be expressed as an $\infty \times \infty$ matrix

$$\tilde{\mathbf{K}} = \frac{1}{(n-p)^2} \tilde{\mathcal{Y}}_0 \sum_{q=1}^{p} \tilde{\mathcal{Y}}_q' \tilde{\mathcal{Y}}_q \tilde{\mathcal{Y}}_0'. \tag{2.29}$$

By letting $\mathbf{A} = \mathcal{Y}_0$ and $\mathbf{B}' = \sum_{1 \le q \le p} \tilde{\mathcal{Y}}_q' \tilde{\mathcal{Y}}_q \tilde{\mathcal{Y}}_0'$, $\mathbf{AB}'$ shares the same nonzero eigenvalues as $\mathbf{B}'\mathbf{A}$. In the other words, $\tilde{\mathbf{K}}$ shares the same nonzero eigenvalues as the $(n-p) \times (n-p)$ matrix

$$\tilde{\mathbf{K}}^* = \frac{1}{(n-p)^2} \sum_{q=1}^{p} \tilde{\mathcal{Y}}_q' \tilde{\mathcal{Y}}_q \tilde{\mathcal{Y}}_0' \tilde{\mathcal{Y}}_0. \tag{2.30}$$

Moreover, let $\tilde{\boldsymbol{\gamma}}_j = (\tilde{\gamma}_{1j}, \ldots, \tilde{\gamma}_{n-p,j})'$ be the eigenvectos of $\tilde{\mathbf{K}}^*$. Then, the eigenfunctions of $\tilde{K}(u, v)$ can be calculated as

$$\sum_{t=1}^{n-p} \tilde{\gamma}_{tj} \{\rho_t(u) - \tilde{\varrho}(u)\}. \tag{2.31}$$

Similarly, we let the curve $\hat{\rho}_t(u) - \hat{\varrho}(u)$ be denoted by the $\infty \times 1$ vector $\hat{\boldsymbol{\varrho}}_t$, from which $\hat{\boldsymbol{\varrho}}_t' \hat{\boldsymbol{\varrho}}_{t+q} = \langle \hat{\rho}_t - \hat{\varrho}, \hat{\rho}_{t+q} - \hat{\varrho} \rangle$ and $\hat{\mathcal{Y}}_q = (\hat{\boldsymbol{\varrho}}_{1+q}, \ldots, \hat{\boldsymbol{\varrho}}_{n-p+q})$. Then, $\hat{K}(u, v)$ can be transformed into an $\infty \times \infty$ matrix

$$\hat{\mathbf{K}} = \frac{1}{(n-p)^2} \hat{\mathcal{Y}}_0 \sum_{q=1}^{p} \hat{\mathcal{Y}}_q' \hat{\mathcal{Y}}_q \hat{\mathcal{Y}}_0', \tag{2.32}$$

which shares the same nonzero eigenvalues as the $(n-p) \times (n-p)$ matrix

$$\hat{\mathbf{K}}^* = \frac{1}{(n-p)^2} \sum_{q=1}^{p} \hat{\mathcal{Y}}_q' \hat{\mathcal{Y}}_q \hat{\mathcal{Y}}_0' \hat{\mathcal{Y}}_0. \tag{2.33}$$

Let $\hat{\theta}_j$ denote a nonzero eigenvalue of $\hat{\mathbf{K}}^*$ and $\hat{\boldsymbol{\gamma}}_j = (\hat{\gamma}_{1j}, \ldots, \hat{\gamma}_{n-p,j})'$ be the corresponding eigenvector, i.e. $\hat{\mathbf{K}}^* \hat{\boldsymbol{\gamma}}_j = \hat{\boldsymbol{\gamma}}_j \hat{\theta}_j$. Then, we are able to compute the eigenfunctions of $\hat{K}(u, v)$ as

$$\sum_{t=1}^{n-p} \hat{\gamma}_{tj} \{\hat{\rho}_t(u) - \hat{\varrho}(u)\}. \tag{2.34}$$

2.4. Theoretical properties

It is important that we first show the uniform convergence rate for the local linear estimator defined in (2.24). Such a uniform convergence is essential in our study since it ensures that the estimated functional correlation is close to the true function everywhere. Assumption 7.1 lists probability and other important time series properties required for all the time series that are involved.

Theorem 2.1. *Let Assumption 7.1 hold. Then we have uniformly:*

$$\hat{\rho}_t(u) = \rho_t(u) + \frac{1}{2} w_2^{\kappa} B_{1\hat{\rho}}(u) h^2 - \frac{1}{2} w_2^{\kappa} B_{2\hat{\rho}}(u) h^2 + N_{\hat{\rho}}(u) + \delta_m, \tag{2.35}$$

where $\delta_m = o_P(h^2 + \{\log m/(mh)\}^{1/2})$,

$$B_{1\hat{\rho}}(u) = \frac{\mu_{k\ell,t}''(u) - \mu_{k,t}(u)\mu_{\ell,t}''(u) - \mu_{\ell,t}(u)\mu_{k,t}''(u)}{\sigma_{\ell,t}(u)\sigma_{k,t}(u)},$$

$$B_{2\hat{\rho}}(u) = \frac{\rho_t(u)(\sigma_{k,t}^2(u))''}{2\sigma_{k,t}^2(u)} + \frac{\rho_t(u)(\sigma_{\ell,t}^2(u))''}{2\sigma_{\ell,t}^2(u)},$$

$$N_{\hat{\rho}}(u) = \frac{1}{mf_{U,t}(u)} \sum_{\tau=1}^{m} \kappa_{h,t,\tau}(u) N_{\hat{\rho},\tau}(u),$$

$$N_{\hat{\rho},\tau}(u) = \frac{e_{k\ell,t,\tau}}{\sigma_{\ell,t}(u)\sigma_{k,t}(u)} - \frac{\rho_t(u)\sigma^2_{k,t}(U_{t,\tau})\xi_{k,t,\tau}}{2\sigma^2_{k,t}(u)} - \frac{\rho_t(u)\sigma^2_{\ell,t}(U_{t,\tau})\xi_{\ell,t,\tau}}{2\sigma^2_{\ell,t}(u)}.$$

$\xi_{k,t,\tau} = \epsilon^2_{k,t,\tau} - 1$, $e_{k\ell,t,\tau} = r_{k,t,\tau} r_{\ell,t,\tau}$ *and* $f_{U,t}(u)$ *is the marginal density of* $U_{t,\tau}$ *whose properties are given in more detail in Assumption 7.1.*

Below let $\{\hat{\psi}_j\}_{j=1}^{\infty}$ denote the eigenfunctions of $\hat{\mathrm{K}}$, for which

$$\begin{aligned} (\hat{\mathrm{K}}\hat{\psi}_j)(u) &= \int_{\mathcal{I}} \hat{K}(u,v)\hat{\psi}_j(v)\, dv \\ &= \frac{1}{(n-p)^2} \sum_{t,s=1}^{n-p} \sum_{q=1}^{p} \{\hat{\rho}_t(u) - \hat{\varrho}(u)\} \langle \hat{\rho}_s - \hat{\varrho}, \hat{\psi}_j \rangle \langle \hat{\rho}_{t+q} - \hat{\varrho}, \hat{\rho}_{s+q} - \hat{\varrho} \rangle \end{aligned} \tag{2.36}$$

and $\hat{\theta}_j$ signifies the corresponding eigenvalue to the eigenfunction $\hat{\psi}_j$. Moreover, let $\|\mathrm{L}\|_{\mathcal{S}}$ denote the Hilbert-Schmidt norm for any operator L (see Appendix 7.1 for detailed definitions). We can now state theoretical properties of $\hat{\mathrm{K}}$, $\hat{\theta}_j$ and $\hat{\psi}_j$. Necessary assumptions and proof are presented in Appendix 7.4.

Theorem 2.2. *Let Assumptions 7.2 hold. Furthermore, let*

$$n = \left\lfloor \left(\frac{m}{\log m} \right)^{4/5} \right\rfloor, \tag{2.37}$$

where $\lfloor Q \rfloor$ *signifies the greatest integer less than or equal to* Q*. Then:*

(i) $\|\hat{\mathrm{K}} - \mathrm{K}\|_{\mathcal{S}} = O_P\left(n^{-1/2}\right)$

(ii) $\sup_{j\geq 1} |\hat{\theta}_j - \theta_j| = O_p(n^{-1/2})$

(iii) $\left[\int_{\mathcal{I}} \{\hat{\psi}_j(u) - \psi_j(u)\}^2\, du\right]^{1/2} = O_P(n^{-1/2})$

In Theorem 2.2, condition (2.37) is given merely as a guideline and for the simplicity of notations. More generally, other combinations of n and m, for example $n \geq m$, are allowed and should only alter the speed of convergence in the theorem. This is also illustrated empirically by simulation results, which are presented in Section 4.

3. Modeling the functional dynamics

For the purposes of correlation analysis and forecasting, it is imperative that we are able to model serial dependence of the FC-TS $\rho_1(u), \ldots, \rho_n(u)$. To achieve such empirical goal, this section employs functional principal components to construct the dynamic space of the curve time series of interest. In other words, we follow a widespread practice in the functional data analysis that is to focus on the truncated expansion in which only d_0 terms is used, namely

$$\mathcal{V}_{d_0,t}(u) = \sum_{j=1}^{d_0} \eta_{tj}\psi_j(u), \quad \eta_{tj} = \int_{\mathcal{I}} \{\rho_t(u) - \mu(u)\}\psi_j(u)du \tag{3.1}$$

(see e.g. Yao et al. (2005), Hall and Hosseini-Nassab (2006), Hall and Vial (2006), Wang (2008), Bathia et al. (2010) and Li et al. (2013)). Such a practice embodies the fact that functional data analysis can be viewed as the functional extension of the principal component analysis. Meanwhile, a parallel assumption is also used regularly in the factor analysis (see e.g. Assumption I1 in Körber et al. (2015) and expression (2.16) of Jiang et al. (2016)).

Moreover, there are a number of results that can help to verify our use of the truncated expansion in (3.1) as an acceptable approximation. Firstly, we have already shown in Lemma 2.1 that the mean squared error using the finite representation in the space of the deterministic function converges to zero. In addition, by using Proposition 1(ii) of Bathia et al. (2010), it holds that

$$\vartheta_{d_0,t}(u) = \sum_{j=1}^{d_0} \xi_{tj}\varphi_j(u) = \mathcal{V}_{d_0,t}(u). \tag{3.2}$$

Using this result, we can also present the optimality of the truncated Karhunen-Loéve expansion as follows:

Lemma 3.1. *Among all truncated expansions expressed in the form of (3.1), the truncated Karhunen-Loéve expansion (3.1) is optimal in the sense that it minimised the integrated mean squared error*

$$\int_{\mathcal{I}} E(e^2_{d_0,t}(u))\ du$$

where $e_{d_0,t}(u) = \sum_{j=d_0+1}^{\infty} \eta_{tj}\psi_j(u)$.

In the sections that follow, we discuss how finite dimensionality is useful in the analysis of the FC-TS.

3.1. *Finite dimensional FC-TS*

Let us begin with the following truncated version of (2.9)

$$M^{(q)}(u,v) = \sum_{i,j=1}^{d_0} \sigma_{ij}^{(q)}\varphi_i(u)\varphi_j(v), \tag{3.3}$$

where $d_0 \geq 1$ and $\mathbf{\Sigma}^{(q)} = E(\boldsymbol{\xi}_t\boldsymbol{\xi}'_{t+q}) \equiv \left(\sigma_{ij}^{(q)}\right)$ is autocovariance matrix of the d_0-dimensional vector process $\boldsymbol{\xi}_t = (\xi_{t1}, \ldots, \xi_{td_0})'$. Under (3.3), the time series evolution of $\vartheta_t(u)$ is driven by that of $\boldsymbol{\xi}_t = (\xi_{t1}, \ldots, \xi_{td_0})'$. Hence, the dynamic (function) space of interest is spanned by the deterministic eigenfunctions $\varphi_1(u), \ldots, \varphi_{d_0}(u)$, namely $\mathcal{M} = \text{span}(\varphi_1(u), \ldots, \varphi_{d_0}(u))$.

Likewise, $N^{(q)}(u,v)$ and $K^{(q)}(u,v)$ (in (2.17) and (2.16) respectively) can be redefined based on the truncation in (3.3). Since the dynamic space $\mathcal{M}$ is now closed, we can show that, for a fixed finite integers $d_0 \geq 1$ and $p \geq 1$, $\hat{\mathcal{M}} = \text{span}(\hat{\psi}_1(u), \ldots, \hat{\psi}_{d_0}(u))$ is a consistent estimator of $\mathcal{M}$. Theorem 3.1 below ensures that, although $\hat{\psi}_j$ are not direct estimators for the eigenfunctions φ_j of $\mathrm{M}^{(0)}$, $\hat{\mathcal{M}} = \text{span}(\hat{\psi}_1(u), \ldots, \hat{\psi}_{d_0}(u))$ is a consistent estimator of the dynamic space $\mathcal{M} = \text{span}(\varphi_1(u), \ldots, \varphi_{d_0}(u))$.

Theorem 3.1. *Let Assumptions 7.2 hold and* $n = \left\lfloor \left(\frac{m}{\log m}\right)^{4/5} \right\rfloor$ *as required in Theorem 2.2. Then, for a given fixed* d_0,

$$D(\hat{\mathcal{M}}, \mathcal{M}) = O_P(n^{-1/2}) \tag{3.4}$$

where $D(\cdot, \cdot)$ *is a discrepancy measure, whose exact definition is given under Definition (v) in Appendix 7.1.*

Theorem 3.1 together with equation (3.2) suggest the fitting

$$\hat{\vartheta}_{d_0,t}(u) = \sum_{j=1}^{d_0} \hat{\eta}_{tj} \hat{\psi}_j(u), \tag{3.5}$$

where $\hat{\eta}_{tj} = \int_{\mathcal{I}} \{\hat{\rho}_t(u) - \hat{\varrho}(u)\} \hat{\psi}_j(u) du$. As the results, to model the dynamic behavior of the FC-TS, we only need to model that of the d_0-dimensional vector process $\widehat{\boldsymbol{\eta}}_t = (\widehat{\eta}_{t1}, \ldots, \widehat{\eta}_{td_0})'$ using one of the many multivariate time series model available in the literature, e.g. the VARMA model.

Remark 3.1. *If* d_0 *is allowed to tend to infinity, we can also obtain the below consistency for* $\hat{\vartheta}_{d_0,t}(u)$. *This result is closely related to that in Lemma 2.1 above.*

Lemma 3.2. *Under the conditions of Theorem 2.2. For* $d_0 \to \infty$ *and* $n \to \infty$, *it holds that*

$$\lim_{d_0 \to \infty} \lim_{n \to \infty} \hat{\vartheta}_{d_0,t}(u) = \vartheta_t(u). \tag{3.6}$$

3.2. Selecting the finite dimensionality, d_0

Under the finite dimensionality of functional time series, it is possible to decompose the space $\mathcal{L}_2(\mathcal{I})$ into $\mathcal{M}$ and $\mathcal{M}^{\perp}$, where $\mathcal{M}^{\perp}$ is the orthonormal complement of $\mathcal{M}$. Since $\mathcal{M}$ is the dynamic space as explained in Section 3.1, $\mathcal{M}^{\perp}$ represents the serially uncorrelated component. In the current section, we construct a class of information criteria for selecting the dimension d_0 (equivalently the number of eigenfunctions spanning the dynamic space $\mathcal{M}$). To do so, we first focus on the basic construction, then explain a few operational issues.

For $1 \leq d \leq d_{\max}$, let

$$\hat{S}^{(d)} = \sum_{j=1}^{d} \langle \hat{\psi}_j, \hat{\mathrm{K}}\hat{\psi}_j \rangle,$$

where $d_{\max}$ denotes a fixed search limit and $(\hat{\mathrm{K}}\hat{\psi}_j)(u)$ as given in (2.36). We suggest the following class of criteria

$$IC(d) = \hat{S}^{(d)} - (d \times P_n), \tag{3.7}$$

where P_n is a penalty function satisfying the conditions stated in Theorem 3.2 below, and identify d_0 as

$$\hat{d} = \max_d IC(d). \tag{3.8}$$

Lemma 3.3 below will be useful for proving the consistency of such a selection.

Lemma 3.3. *Let Assumptions 7.2 hold and* $n = \left\lfloor \left(\frac{m}{\log m}\right)^{4/5} \right\rfloor$ *as in Theorem 2.2. Furthermore, let* $\sum_{j=1}^{d_0} \hat{\theta}_j = \sum_{j=1}^{d_0} \langle \psi_j, \hat{\mathrm{K}}\hat{\psi}_j \rangle$ *and* $\sum_{j=1}^{d_0} \theta_j = \sum_{j=1}^{d_0} \langle \psi_j, \mathrm{K}\psi_j \rangle$*. Then, as* $n \to \infty$,

$$\sum_{j=1}^{d_0} (\hat{\theta}_j - \theta_j) = O_P(n^{-1/2}) \quad \text{and} \quad \sum_{j=d_0+1}^{n} \hat{\theta}_j = O_P(n^{-1}). \tag{3.9}$$

These results relate closely to $\sum_{j=1}^{\infty} \langle \varphi_j, \mathrm{M}^{(0)}\varphi_j \rangle = \sum_{j=1}^{\infty} \lambda_j$, which describes the total covariance in the traditional functional data analysis. In the context of this paper, $\sum_{j=1}^{\infty} \theta_j$ signifies the total auto-covariance in the functional time series in question, so that $\sum_{j=1}^{d_0} \theta_j / \sum_{j=1}^{\infty} \theta_j$ quantifies the proportion of the total auto-covariance explained by the d_0-truncation.

Theorem 3.2. *Let Assumptions 7.2 hold and* $n = \left\lfloor \left(\frac{m}{\log m}\right)^{4/5} \right\rfloor$ *as required in Theorem 2.2. Suppose that the penalty function* P_n *satisfies (a)* $P_n \to 0$*, and (b)* $C_n P_n > 1$ *for* $n \to \infty$*, where* $C_n = n^{1/2}$.

(i) Let $\hat{d}$ *be the maximiser of the information criteria among* $1 \leq d \leq d_{\max}$*, where* $d_{\max}$ *denotes a fixed search limit. Then:*

$$\lim_{n \to \infty} Prob(\hat{d} = d_0) = 1 \tag{3.10}$$

(ii) The consistency in (3.10) still holds for the case where $d_0 = d_n$ *is considered a function of* n *and tends to infinity more slowly than* $n^{1/2}$.

Under the conditions of the theorem, Theorem 3.2(i) confirms that $\hat{d}$ selected based on (3.8) is a consistent estimator of d_0. While Lemma 3.2 implies that we must also consider the case in which $d_0 = d_n$, where d_n is a function of sample size n, tending to infinity in order to maintain the consistency of the representation, Theorem 3.2(ii) shows that theoretically $\hat{d}$ selected based on

Table 1

Percentages of accurate dimension selection across the (m, n)-pairs and simulation repetitions

m,n	16	45	60	80	114	200	300	400
75	36.5	50.5	56.5	58.0	69.5	77.5	86.5	91.5
390	35.0	71.5	69.5	82.0	90.0	97.5	97.5	100.0
600	24.0	63.0	75.5	81.0	59.0	98.5	100.0	100.0
1000	43.0	63.0	75.0	78.0	86.0	98.0	100.0	100.0
1600	35.0	63.0	76.0	85.0	90.0	100.0	100.0	100.0

(3.2) does also comply with such a tendency. It is required that d_n must tend to infinity more slowly than $n^{1/2}$, however. In this regard, it is consistent with the results of the theorem to set $d_{\max} = n/A$ for some $A > 1$ (e.g. $d_{\max} = n/\log n$) since

$$\sum_{j=1}^{n} \hat{\theta}_j = \sum_{j=1}^{\infty} \langle \psi_j, \hat{\mathrm{K}} \hat{\psi}_j \rangle$$

due to the Eigendecomposition $\hat{\mathbf{K}}^* \hat{\boldsymbol{\gamma}}_j = \hat{\boldsymbol{\gamma}}_j \hat{\theta}_j$ in Section 2.3.

In the context of the factor analysis, Bai and Ng (2002) propose a class of information criteria whereby the penalty term shows symmetry in the roles of m and n. In this paper, we apply the local-linear estimators along m, and hence m and n play different roles in our rate. The information criteria that satisfy conditions (a) and (b) in Theorem 3.2 can be constructed as follows

$$IC_1(d) = \hat{S}^{(d)} - \left(d \times \left\{ \frac{\log n}{n} \right\}^{\nu_1} \right), \quad \nu_1 = \left\lfloor \frac{1}{2} \left\{ \frac{\log n}{\log\, (n/\log n)} \right\} \right\rfloor$$

and

$$IC_2(d) = \hat{S}^{(d)} - \left(d \times B^{\nu_2} \right), \ \nu_2 = \left\lfloor \frac{1}{2} \left\{ \frac{\log B}{\log\, (B/\log B)} \right\} \right\rfloor ,$$

where $B = \left(\frac{n+m}{nm}\right) \log \left(\frac{nm}{n+m} \right)$.

4. Simulation studies

In this section, we conduct a number of simulation exercises. In doing so, we are interested in examining the finite sample performance of (a) the information criteria $IC(d)$ for selecting the number of eigenfunctions d_0 that span the dynamic space $\mathcal{M} = \mathrm{span}(\varphi_1, \ldots, \varphi_{d_0})$, (b) the estimator $\hat{\mathcal{M}} = \mathrm{span}(\hat{\psi}_1, \ldots, \hat{\psi}_{d_0})$ as an estimator of the dynamic space $\mathcal{M}$ and (c) the local linear estimator $\hat{\rho}_t(u)$. Let us begin with $IC(d)$ and $\hat{\mathcal{M}}$ as follows.

4.1. Finite sample performance of $IC(d)$ and $\hat{\mathcal{M}}$

To this end, we consider again the pair of asset returns that were defined just above equation (2.1). In this regard the correlation coefficient defined in (2.1)

TABLE 2
Medians of the D measure (defined in (4.3)) across (m,n)-pairs and repetitions

m/n	16	45	60	80	114	200	300
75	0.5172	0.4810	0.4463	0.3845	0.3597	0.3314	0.3092
390	0.2813	0.2219	0.2031	0.1985	0.1851	0.1696	0.1523
600	0.2541	0.1927	0.2030	0.1653	0.1505	0.1423	0.1352
1000	0.2016	0.1805	0.1453	0.1383	0.1307	0.1192	0.1162
1600	0.1641	0.1398	0.1281	0.1260	0.1117	0.1204	0.1022

is in fact $E[\epsilon_{k,t,\tau}\epsilon_{\ell,t,\tau}|U_{t,\tau}]$. Hence, we are able to generate as a model example

$$\epsilon_{k,t,\tau}\epsilon_{\ell,t,\tau} = \varrho_{\epsilon,t}(U_{t,\tau}) + e_{t,\tau},\ e_{t,\tau} \overset{i.i.d.}{\sim} N(0,1),\ \tau = 1,\dots,m,$$

where $\varrho_{\epsilon,t}(U_{t,\tau}) \equiv E[\epsilon_{k,t,\tau}\epsilon_{\ell,t,\tau}|U_{t,\tau}]$. We shall assume in this section that

$$\begin{aligned} E[\epsilon_{k,t,\tau}\epsilon_{\ell,t,\tau}|U_{t,\tau}=u] &= \sum_{i=1}^{d_0}\xi_{it}\varphi_i(u) + \sum_{j=1}^{10}\frac{Z_{jt}}{2^{j-1}}\zeta_j(u), \\ &\equiv \varrho_{\epsilon,t}(u), \quad u\in[0,1] \end{aligned} \tag{4.1}$$

where $\varphi_i(u) = \sqrt{2}\cos(\pi i u)$ with loading series $\{\xi_{it}, t \geq 1\}$ following a linear AR(1) process with coefficient $(-1)^i(0.9-0.5i/2)$ and $\zeta_j(u) = \sqrt{2}\sin(\pi j u)$ whereas Z_{jt} are independent $N(0,1)$ variables. We then treat (4.1) as a correlation trajectory that is assumed to be a realisation of the functional correlation time series of interest.

In this regard, the empirical estimation begins with constructing

$$\hat\varrho_{\epsilon,t}(u) = \frac{\sum_{\tau=1}^m W_{m,h}(U_{t,\tau}-u)\epsilon_{k,t,\tau}\epsilon_{\ell,t,\tau}}{\sum_{\tau=1}^m W_{m,h}(U_{t,\tau}-u)},\quad t=1,\dots,n, \tag{4.2}$$

where $W_{m,h}(U_{t,\tau}-u) = s_{m,h,2}\kappa_h(U_{t,\tau}-u) - s_{m,h,1}\kappa_h(U_{t,\tau}-u)(U_{t,\tau}-u)$, $s_{m,h,j} = \sum_{\tau=1}^m \kappa_h(U_{t,\tau}-u)(U_{t,\tau}-u)^j$ for $j=0,1,2$, $\kappa_h(U_{t,\tau}-u) = \kappa\big(\frac{U_{t,\tau}-u}{h}\big)/h$ and $\kappa(\cdot)$ is a kernel function. h is the bandwidth parameter, which in practice is selected based on the cross-validation method. We then use the functional process $\hat\varrho_{\epsilon,1}(u),\dots,\hat\varrho_{\epsilon,n}(u)$ in place of $\varrho_{\epsilon,1}(u),\dots,\varrho_{\epsilon,n}(u)$ when selecting the number of eigenfunctions $\hat d$ and computing $\hat{\mathcal{M}} = \mathrm{span}(\hat\psi_1,\dots,\hat\psi_{d_0})$. Statistical validity of the above-discussed set-up for checking the finite sample performance of interest is ensured by noting that uniformly

$$\hat\varrho_{\epsilon,t}(u) - \varrho_{\epsilon,t}(u) = \frac{1}{2}w_2^\kappa\varrho''_{\epsilon,t}(u)h^2 + \frac{1}{mf_{U,t}(u)}\sum_{\tau=1}^m \kappa_{h,t,\tau}(u)e_{t,\tau} + \delta_m,$$

where $\kappa_{h,t,\tau}(u) \equiv \kappa_h(U_{t,\tau}-u)$ and $\delta_m = o_P(h_t^2 + \{\log\ m/(mh)\}^{1/2})$, which was established in the proof of Theorem 3.1 of Jiang et al (2015). Such a result is in line with the uniform convergence rate shown in our Theorem 2.1.

Moreover, we measure the discrepancy between $\hat{\mathcal{M}} = \text{span}(\hat{\psi}_1, \ldots, \hat{\psi}_{d_0})$ and the dynamic space $\mathcal{M} = \text{span}(\varphi_1, \ldots, \varphi_{d_0})$ by the metric

$$D(\hat{\mathcal{M}}, \mathcal{M}) = \sqrt{1 - \frac{1}{d_0} \sum_{j,k=1}^{d_0} (\langle \hat{\psi}_j, \varphi_k \rangle)^2}, \tag{4.3}$$

where

$$\sum_{j,k=1}^{d_0} (\langle \hat{\psi}_j, \varphi_k \rangle)^2 \leq 1,$$

which suggests that D is a symmetric measure between 0 and 1.

To conduct our simulation exercises, we set $d_0 = 2$, so that the dynamics of the functional time series is driven only by that of ξ_{1t} and ξ_{2t}. In addition, let $d_{max} = 5$ and $p = 5$. The exercises are conducted under 200 simulation repetitions and results are compared among various combinations of m and n, which are shown by the rows and columns of Table 1. Quantities presented in the table are the percentages of correct selection made based on $IC(d)$. Overall, it is clear that an increase in either m- or n-direction improves the accuracy of the dimension selection. In addition, at $m = 390$ the best possible outcome of 100% accuracy is achieved at $n = 400$, while it is achieved at only $n = 300$ when $m \geq 600$. Nonetheless, Figure 1 shows some evidence that improvement in the performance tails off when n increases beyond the relative magnitude recommended as a condition of Theorem 3.2. The most convenient way to perceive this is to recognize the curvature of the graphs with declining (positive) slope as n increases. Let us take as an example the case where $m = 390$. Here the percentage increases sharply as $n = 16$ increases to $n = 45$, but the improvement is at much slower rate when n is increased beyond this point. A similar argument is also applicable to other values of m.

We now investigate how effective $\hat{\mathcal{M}} = \text{span}(\hat{\psi}_1, \hat{\psi}_2)$ is in finite sample as an estimator of the dynamic space $\mathcal{M} = \text{span}(\varphi_1, \varphi_2)$. Table 2 presents medians of the D measure defined in (4.3) across the (m, n)-settings. Overall, it can be concluded that an increase in either m or n leads to more accurate estimation of the dynamic functional space. However, Figure 2 shows some evidence that the improvement tails off when n increases beyond the relative magnitude recommended as a condition of Theorem 3.1. The most convenient way to establish this is to recognize the curvature of the graphs with declining (negative) slope as n increases. Let us take the case where $m = 390$ as an example. The drop of the median when $n = 16$ increases to $n = 45$, which is the recommended rate, is much sharper than other ones. Another example is when $m = 600$ when the rate of improvement declines as n increases beyond 60. A similar phenomenon is seen across all values of m. These provide empirical evidence in support of our argument that the asymptotic rates of functional time series analysis are affected by the estimation of correlation functions in question when n is beyond what recommended by the (m, n)-relation.

TABLE 3

Finite sample performance comparison: Our local linear (LL) versus Hafner's et al. (2006) local constant (LC) estimators of correlation function

		ASE_{Cov}		ASE_S	
		LL	LC	LL	LC
	75	2.3872e-03	2.3996e-03	1.6938e-03	1.5022e-03
	390	1.8404e-03	2.0811e-03	4.1834e-04	4.4751e-03
m	600	1.7630e-03	1.9861e-03	3.09904-04	3.4533e-04
	1000	1.6092e-03	1.7708e-03	2.0121e-04	2.3924e-03
	1600	1.5339e-03	1.6964e-03	1.7437e-04	2.0108e-04

4.2. Finite sample performance of $\hat{\rho}_t(u)$

In this regard, the key motivation is to ensure that the estimated functional correlation is close to the true function everywhere. In addition, we shall also compare the finite sample performance our local linear estimator, $\hat{\rho}_t(u)$, to that of the SP-C of Hafner et al. (2006). To this end, we assume that the return process follows

$$r_{j,t,\tau} = a_{jt} + b_{jt}\mu_{j,t}(U_{t,\tau}) + c_{j0,t}\epsilon_{t,\tau} + c_{j1,t}f_1(U_{t,\tau})\epsilon_{1,t,\tau} + c_{j2,t}f_2(U_{t,\tau})\epsilon_{2,t,\tau},$$

where a_{jt}, b_{jt}, $c_{j0,t}$, $c_{j1,t}$, $c_{j2,t}$ are constant coefficients, $j = k, \ell$, and $\epsilon_{0,t,\tau}$, $\epsilon_{1,t,\tau}$, $\epsilon_{2,t,\tau}$, are random renovations with zero mean. We also assume $\mu_{j,t}(U_{t,\tau}) = U_{t,\tau}$, $a_{jt}, b_{jt}, c_{j0,t}, c_{j1,t}, c_{j2,t} \sim N(0, 0.2)$, $\epsilon_{0,t,\tau}, \epsilon_{1,t,\tau}, \epsilon_{2,t,\tau} \sim N(0, 1)$,

$$f_1(U_{t,\tau}) = \sqrt{1 + \cos(2\pi U_{t,\tau})} \quad \text{and} \quad f_2(U_{t,\tau}) = \sqrt{1 + \sin(2\pi U_{t,\tau})}.$$

The correlation coefficient of the above returns can then be derived as

$$Corr_t(r_{k,t,\tau}, r_{\ell,t,\tau}|U_{t,\tau} = u) = \frac{Cov_t(r_{k,t,\tau}, r_{\ell,t,\tau}|U_{t,\tau} = u)}{S_t(u)}, \tag{4.4}$$

where $Cov_t(r_{k,t,\tau}, r_{\ell,t,\tau}|U_{t,\tau} = u) = \alpha_t + \beta_t f_1^2(u) + \gamma_t f_2^2(u) \equiv Cov_t(u)$, $\beta_t = c_{k1,t}c_{\ell 1,t}$, $\alpha_t = c_{k0,t}c_{\ell 0,t} + c_{k1,t}c_{\ell 1,t} + c_{k2,t}c_{\ell 2,t}$, $\gamma_t = c_{k2,t}c_{\ell 2,t}$ and $S_t(u) = \sqrt{\sigma^2_{k,t}(u)\sigma^2_{k,t}(u)}$.

To examine the finite sample performance of the estimators in questions, we consider the following measures of discrepancy:

$$ASE_{Cov} = \frac{1}{m}\sum_{\tau=1}^{m}\{\hat{C}ov_t(U_{t,\tau}) - Cov_t(U_{t,\tau})\}^2 \tag{4.5}$$

$$ASE_S = \frac{1}{m}\sum_{\tau=1}^{m}\{\hat{S}(U_{t,\tau}) - S_t(U_{t,\tau})\}^2 \tag{4.6}$$

Our local-linear and Hafner et al (2006) local-constant estimators are referred to in Table 3 as "Local Linear" (LL) and "Local Constant" (LC), respectively.

Table 4
Table of Abbreviations

Abbreviations	Definitions
JPY	Japanese Yen
EUR	European Union Euro
USD	United States Dollar
CHF	Swiss Franc
GBP	British Pound
NOK	Norwegian Krone
SEK	Swedish Krona
jpy	USD/JPY Exchange rate
eur	USD/EUR Exchange rate
chf	USD/CHF Exchange rate
gbp	USD/GBP Exchange rate
nok	USD/NOK Exchange rate
sek	USD/SEK Exchange rate
$\rho_{chf,t}(u)$	Correlation process between gbp and chf returns
$\rho_{nok,t}(u)$	Correlation process between gbp and nok returns
$\rho_{sek,t}(u)$	Correlation process between gbp and sek returns

Although the simulation results in Table 3 suggests that both estimators perform well in finite sample, our local linear estimator seems to have a clear edge on its local constant counterpart. An intensive graphical examination suggests that the local linear estimator enjoy better performance near the boundary as ones can expect.

5. Empirical Analysis of Exchange Rate Returns and Correlations

Table 4 presents a list of abbreviations used in the current section. Let us begin with a brief motivation.

5.1. Overview and motivation

In this section, we intend to study co-movements between three pairs of exchange rate returns, namely (i) gbp and chf; (ii) gbp and nok; and (iii) gbp and sek. This study is interesting due to the fact that the UK, Switzerland, Norway and Sweden are large trading partners of each other. Moreover, they share an important characteristic of being a small open economy with a large international financial sector.

Even though Van Dijk et al. (2006) studied such co-movements previously based on the DCC-GARCH model, economic theory has connected exchange rates movements to a large number of macroeconomic factors. A candidate list of economic variables, which can potentially be key drivers of exchange rate returns correlations, is clearly very large so much so that searching over all possibilities might be infeasible. In contrast to this more traditional treatment, Verdelhan (2018) found that the evolution of exchange rates through time can be quite successfully explained by a small number of latent common factors. These factors remained significant and were quantitatively important even after controlling for macroeconomic fundamental determinants of exchange rates (see also Engel et. al. (2015)). Similarly, Greenaway-McGrevy et. al. (2015) formulated three most significant common factors, which drove co-movements of a panel of 27 USD-based exchange rates in their study, and were able to established these factors as the empirical counterparts of the eur, chf and jpy. Due to the eur and jpy domination in foreign exchange trading and the safe-haven role of the jpy and chf, such identification seems economically reasonable.

The objective of the empirical study in this section is to extend the work of Van Dijk et. al. (2006) to studying the time series properties of the FC-TS for (i) gbp and chf returns, (ii) gbp and nok returns and (iii) gbp and sek returns. We make use of the knowledge provided by Greenaway-McGrevy et. al. (2015) and Verdelhan (2018) and treat eur as the driver of the exchange rate return correlations. Below, let us begin with calculation of the returns series and their devolatilization.

5.2. Returns series and devolatilization

The data used in our study are regular interval exchange rate spot prices at 1-minute interval provided by Olsen Data between 1 January 2016 to 30 June 2017. For our dataset, we have found that the majority of the trades fall between midnight and 07:30PM each weekday and therefore excluded weekends and the periods of weekdays outside of these hours. We have also excluded Christmas and New Year holidays, which are 24 to 26 and 31 December 2016, and 1 to 2 January 2017. By letting $p_{j,t,\tau}$ denote the τ intraday spot price of the j exchange rate in the t day, then one-minute returns are computed as $100 \times \log \left\{ \frac{p_{j,t,\tau}}{p_{j,t,\tau-1}} \right\}$, where j denotes either eur, chf, gbp, dkk, nok or sek. These data arrangements and calculations lead to $m = 1,185$ one-minute returns in each of the $n = 388$ days. Moreover, to encourage autoregressive homoscedasticity, we compute devolatilized returns, whereby the devolatilization is performed based on the ARMA(1,1)+GARCH(1,1) process. Then, these devolatilized returns are used in the local-linear estimation, from which the resulting estimates are treated as correlation trajectories that are assumed to be realisations of the functional correlation time series of interest.

5.3. Model estimation and fitting

Our analysis in this section aims to achieve two objectives as follows. Firstly, it is to compute the fitting

$$\hat{\rho}_{k,t}^{(\hat{d}_k)}(u) = \hat{\varrho}_k(u) + \sum_{j=1}^{\hat{d}_k} \hat{\eta}_{k,t,j} \hat{\psi}_{k,j}(u), \tag{5.1}$$

where $\hat{\varrho}_k(u)$ is the estimate of the mean function, k is either chf, nok or sek, and $\hat{d}_k$ is the number of eigenfunctions selected using the information criteria discussed in Section 3.2. Secondly, it is to evaluate how well this approximation is able to capture time series evolution of the FC-TS in question.

As pointed out in the previous section, we are interested in studying co-movements between three pairs of exchange rate returns, namely (i) *gbp* and *chf*, (ii) *gbp* and *nok*, and (iii) *gbp* and *sek*. To keep our discussion organised, in what follow we shall focus on each of these pairs in a separate section. However, since our analysis of the first pair provides an analytical structure for those that follow, it will be discussed in more detail.

5.3.1. *Correlation analysis for the gbp & chf returns*

Regarding the first objective, we shall present our results and discussion in four steps as follows:

Step 5.1: Firstly, it is the local-linear estimation of daily correlation $\rho_{chf,t}(u)$. Figure 3 presents the 2-dimension and 3-dimension plots of

$$\hat{\rho}_{chf,1}(u_{t,\tau}), \ldots, \hat{\rho}_{chf,n}(u),$$

which are estimated FC-TS for the *gbp* and *chf* returns. In the panel (b) of the figure, $\hat{\rho}_{chf,1}(u)$ is also drawn in the blue color as an example. Since various local-linear estimators are needed in the production of these estimates, a single theoretically-optimal bandwidth, namely $\{\log m/m\}^{1/5}$, is used.

Step 5.2: The second step involves estimating the mean correlation function, $\varrho_{chf}(u)$. This is done based on

$$\hat{\varrho}_{chf}(u) = \frac{1}{n} \sum_{1 \leq j \leq n} \hat{\rho}_{chf,j}(u), \tag{5.2}$$

which is analogous to that in (2.27). Figure 3 presents $\hat{\varrho}_{chf}(u)$ as a (right-scaled) thick blue curve in its top panel. This shows that correlations between the *gbp* and *chf* returns are higher at both ends of the *eur* return spectrum. In addition to (5.2), we compute an alternative estimate based on the formula in (2.24) by using the data across $\tau = 1, \ldots, m$ and $t = 1, \ldots, n$. This is methodologically comparable to the semiparametric estimator introduced in Hafner et al. (2006) and leads to a correlation trajectory, which shares similar features to that presented in Figure 3.

Step 5.3: The third step involves using the above-introduced information criteria to select $\hat{d}_{chf}$. In doing so, we set the maximum search limit at $d_{\max} = 10$. Figures 4 presents $IC_{chf}(d)$ scores, which suggest that

$$\hat{d}_{chf} = \max_{d} IC_{chf}(d) = 5.$$

It is important to note these scores are computed based on $IC_1(d)$, while the use of $IC_2(d)$ also results in a similar selection. In addition, such a selection is congruent with evidence we obtain from the autocorrelation functions (ACFs) of the time series of loadings $\hat{\eta}_{chf,t,1}, \ldots, \hat{\eta}_{chf,t,6}$, which are presented in Figure 5.

The ACFs of $\hat{\eta}_{chf,t,j}$ show much weaker evidence of serial correlation for $j \geq 5$. Finally, Figure 6 presents estimated of the eigenfunctions corresponding to the first five nonzero eigenvalues, i.e. $\hat{\psi}_{chf,1}(u), \ldots, \hat{\psi}_{chf,5}(u)$.

Step 5.4: By using the results of Steps 5.1 to 5.3, we can now compute $\hat{\rho}^{(5)}_{chf,t}(u)$, which can be treated as in-sample forecasts for $\rho_{chf,t}(u)$. Figure 7 presents $\hat{\rho}^{(5)}_{chf,t}(u)$ (black), $\hat{\rho}_{chf,t}(u)$ (red) and $\hat{\varrho}_{chf}(u)$ (blue) for eight randomly selected days. Overall, the predictions are reasonably close to the consistent estimated of the daily realized correlation functions.

We shall now focus on the second objective, i.e. to examine how well the functional process $\hat{\rho}^{(5)}_{chf,t}(u)$ can capture serial correlation in the functional time series $\rho_{chf,1}(u), \ldots, \rho_{chf,n}(u)$. We answer this question in three steps as follows.

Firstly, analogous to a case of the traditional functional data analysis, here we construct a measure

$$PAE\left(\hat{d}_{chf}\right) = \sum_{d=1}^{\hat{d}_{chf}} \hat{\theta}_{chf,d} / \left(\sum_{j=1}^{n} \hat{\theta}_{chf,j}\right). \tag{5.3}$$

In accordance with Theorem 3.3, this should help to quantify the percentage of autocovariance of the time varying component being explained. In fact, such a measure can be computed over $1 \leq d \leq d_{\max} = 10$ as shown in Figure 8. The figure shows that up to 99.03% of autocovariance is explained at $\hat{d}_{chf} = 5$.

Secondly, we compare our in-sample forecasts to those based on the SP-C model of Hafner et al. (2006). Recall firstly that by setting the time-varying component of the correlation to zero, the time-invariant part of our model, i.e. $\varrho_{chf}(u)$, is analogous to an estimate ones can obtain using method introduced in Hafner et al. (2006) (see also discussion in Section 2.1 and in Step 5.2 above). In this regard, the results in Figure 7 does provide some useful information. Taking a role of an in-sample forecast, $\hat{\rho}^{(5)}_{chf,t}(u)$ clearly do reasonably well in predicting the correlation trajectories for the eight randomly selected t. On the contrary, $\hat{\varrho}_{chf}(u)$ is as accurate only around the zero *eur* return. In the figure, the differences between the black and blue trajectories becomes larger as we move further to both extreme ends of the *eur* returns spectrum.

Finally, it should also be useful to compare the performance of our method to that of the DCC-GARCH model. Such comparison should be most meaningful when performed based on $\hat{\rho}^{(5)}_{chf,t}(u)$ and correlation forecasts based on the DCC-GARCH at the daily frequency. However, having based our model and its estimation on one-minute returns means that such a procedure could involve a high degree of uncertainty. As an alternative approach, we shall concentrate instead on contrasting the types of time series evolution enabled in our method against the GARCH-type dynamics specified in the DCC-GARCH. Following the functional time series approach, the dynamics of the FC-TS is driven by that of the loading time series $\eta_{chf,1,t}, \ldots, \eta_{chf,5,t}$. Since the first three eigenfunctions can already explain more than 96% of the total autocovariance (as indicated in Figure 8), we will only focus on $\hat{\eta}_{chf,1,t}, \hat{\eta}_{chf,2,t}$ and $\hat{\eta}_{chf,3,t}$. In Figure 5, the ACFs of $\hat{\eta}_{chg,t,1}$ expresses a strong degree of persistence, while those of $\hat{\eta}_{chf,t,2}$ suggest presence of some cyclical behavior. A careful look at the plots reveals that the former may be caused by some low frequency cycles with relatively lengthy periodicity and trend, while the latter is caused by high frequency cycles (say, for example, the day-of-the-week effects) with a shorter periodicity. Clearly, the GARCH-type dynamics specified in the DCC-GARCH is not able to capture these features. In this regard, the nonparametric method introduced by Aslanidis and Casas (2013) should be more effective in capturing these features.

5.3.2. Correlation analysis for (i) gbp & nok, and (ii) gbp & sek returns

The discussion in this section closely follow the analytical structure used in Section 5.3.1. Let us discuss some important findings below.

Figure 9 presents 2- and 3-dimension plots of $\hat{\rho}_{sek,1}(u), \ldots, \hat{\rho}_{sek,n}(u)$, which are the FC-TS for the *gbp* and *sek* returns. Panel (b) of the figure also presents $\hat{\rho}_{sek,1}(u)$ in the blue color as an example. On other hand, those estimates for the *gbp* and *nok* returns are presented in Figure 15. Judging from the color of the surface plot, overall the FC-TS computed for the *gbp* and *sek* returns appears to display weaker serial correlation compared to those of the remaining pairs.

Figure 9(a) and 15(a) presents, as the dark blue curves, estimates of the expected correlations $\hat{\varrho}_{sek}(u)$ and $\hat{\varrho}_{nok}(u)$, respectively. These estimates represent the time-invariant part, show that correlations between the *gbp* returns and those of *nok* and *sek* are higher at both ends of the *eur* return spectrum. In addition, there exists clear evidence of asymmetry in the effects of the *eur* return on the exchange rate return correlations.

Figures 10 and 16 present the IC scores, $IC_{sek}(d)$ and $IC_{nok}(d)$, respectively. These figures show that

$$\hat{d}_{sek} = \max_d IC_{sek}(d) = 4 \quad \text{and} \quad \hat{d}_{nok} = \max_d IC_{nok}(d) = 5.$$

These results are similar to that presented for the *gbp* and chf returns correlation and indeed congruent with the autocorrelation functions presented in Figures 11 and 17. It is quite noticeable, however, that the autocorrelation function associated with $\hat{\eta}_{nok,t}$ shown a high degree of persistence.

Figures 12 and 18 presents the estimated eigenfunctions, $\hat{\psi}_{sek,1}, \ldots, \hat{\psi}_{sek,6}$, and $\hat{\psi}_{nok,1}, \ldots, \hat{\psi}_{nok,6}$, respectively. These correspond to the first five largest eigenvalues. Overall the shape of the first to forth eigenfunctions appears to be quite similar across the three pairs of returns under consideration. However, those based on the FC-TS of *gbp* and *sek* returns seem to display much stronger degree of curvature.

Figures 13 compares the fittings $\hat{\rho}^{(4)}_{sek,t}(u)$, which represent the in-sample forecasts, to the estimates $\hat{\rho}_{sek,t}(u)$ and those of the non-time-varying parts. Clearly, $\hat{\rho}^{(4)}_{sek,t}(u_{t,\tau})$ do reasonably well in predicting the correlation trajectories for the eight randomly selected t. An analogous comparison between $\hat{\rho}^{(5)}_{nok,t}(u)$

and $\hat{\rho}_{nok,t}(u)$ is presented in Figure 19 and draws a similar set of findings. However, the performance of the time-invariant part as an in-sample forecaster seems to worsen.

Figures 14 and 20 present the percentage autocovariance of the FC-TS of *gbp* and *sek* returns, and *gbp* and *nok* returns being explained, respectively. Although the plot of the latter is closely similar to the previous case in Section 5.3.1, that of the former displays some peculiar features. Figure 14 shows that less 90% of the autocovaraince is explained by the first three functional principal components, compared to just below 97% and 99% for cases of chf and nok, respectively.

6. Conclusions

We studied an alternative approach for modeling time varying behavior of asset returns co-movements. To do so, we took the view that co-movements between a pair of asset returns could be explained entirely by a trajectory of the returns' correlation. The time-series evolution and serial dependence of such trajectories were captured by a functional process that was constructed as a combination of a time-invariant and a time-varying components. The resulting procedure was not only able to address previous limitations of FDA in financial applications, but also offered a general specification that is able to model processes of time-varying time-series correlations generated under many existing models. For practical purpose, our approach treated the correlation coefficient of asset returns for each day as a correlation trajectory that was assumed to be a realisation of the functional time series of interest. Hence, our procedure began with the local-linear estimation of the correlation coefficient in question, which then led to construction of the linear operator based on an auto-covariance kernel. Subsequently, solving for the relevant eigenvalues and eigenfunctions are performed by transforming the problem into an eigenanalysis for a finite matrix. Moreover, our approach relied on functional principal components in our construction of the dynamic space for the functional correlation time series of interest. In this paper, we established a new class of information criteria to help to identify the finite dimensionality of the curve time series. To verify the use of the truncated expansion as a reasonable approximation, we established both consistency and optimality of such a representation. We also established a set of asymptotic results in order to show the statistical validity of the proposed estimation procedure. To illustrate its empirical relevance, we conducted a series of simulation studies and applied our analytical framework to model time varying correlation of exchange rate returns for a group of small open economies with large financial sectors. Our empirical results indicated that concepts of time varying correlation enabled by existing methods, especially the SP-C and the DCC-GARCH models, are too restrictive to accommodate fully the time-varying behavior and temporal evolution of the returns correlation. Finally, our empirical results suggested that the time series evolution of returns correlation involved both low frequency cycles with relatively lengthy periodicity and trend, and high frequency cycles (say, for example, the day-of-the-week effects) with a shorter periodicity.

7. Appendix

7.1. Definitions

The below definitions will be useful in the discussion that follows.

(i) Let $\mathcal{H}$ be a real separable Hilbert space with respect to some inner product $\langle\cdot,\cdot\rangle$. Also, let L be a linear operator from $\mathcal{H}$ to $\mathcal{H}$. For $x \in \mathcal{H}$, let us denote by Lx the image of x under L. In addition, the adjoint of L is denoted by L* and satisfies

$$\langle \mathrm{L}x, y\rangle = \langle x, \mathrm{L}^* y\rangle, \quad x, y \in \mathcal{H}.$$

Accordingly, L is said to be self adjoint if L* = L and nonnegative definite if

$$\langle \mathrm{L}x, x\rangle \geq 0 \quad \forall x \in \mathcal{H}.$$

(ii) For a real separable Hilbert space, e.g. $\mathcal{H}$, let $\|\cdot\|$ denote norm generated by an inner product $\langle\cdot,\cdot\rangle$. Let $\mathcal{B} = \mathcal{B}(\mathcal{H},\mathcal{H})$ denote the space of bounded linear operators form $\mathcal{H}$ to $\mathcal{H}$.

(iii) When $\mathcal{H} = \mathcal{L}_2(\mathcal{I})$ equipped with the inner product defined in (2.3), a compact operator $\mathrm{L} \in \mathcal{B}$ is defined as $(\mathrm{L}x)(u) = \int_{\mathcal{I}} L(u,v)x(v)dv$. In addition, if there exists two orthonormal sequences $\{e_j\}$ and $\{f_j\}$ of $\mathcal{H}$, and a sequence of scalars $\{\lambda_j\}$ decreasing to zero, then

$$(\mathrm{L}x)(u) = \sum_{j=1}^{\infty} \lambda_j \langle e_j, x\rangle f_j(u).$$

(iv) The Hilbert-Schmidt norm of the compact linear operator L is defined as

$$\|\mathrm{L}\|_{\mathcal{S}} = \left(\sum_{j=1}^{\infty} \lambda_j^2\right)^{1/2}.$$

In addition, let $\mathcal{S}$ denote the space consisting of all the operators with a finite Hilbert-Schmidt norm.

(v) Let $\mathcal{N}_1$ and $\mathcal{N}_2$ be any two d_0-dimensional subspaces of $\mathcal{L}_2(\mathcal{I})$, where $\mathcal{L}_2(\mathcal{I})$ denotes the Hilbert space consisting of all the square integrable curves defined on $\mathcal{I}$. In addition, let $\{\zeta_{i1}(\cdot),\ldots,\zeta_{id_0}(\cdot)\}$ be an orthonormal basis of $\mathcal{N}_i$, $i = 1, 2$. Then the projection of ζ_{1k} onto $\mathcal{N}_2$ may be expressed as $\sum_{j=1}^{d_0}\langle\zeta_{2j},\zeta_{1k}\rangle\zeta_{2j}(u)$, $u \in \mathcal{I}$, while the discrepancy between $\mathcal{N}_1$ and $\mathcal{N}_2$ is measured by

$$D(\mathcal{N}_1,\mathcal{N}_2) = \sqrt{1 - \frac{1}{d_0}\sum_{j,k=1}^{d_0}(\langle\zeta_{2j},\zeta_{1k}\rangle)^2}. \tag{7.1}$$

(vi) Let $\mathcal{Z}$ be the set consisting of all the d_0-dimensional subspaces in $\mathcal{L}_2(\mathcal{I})$. Then $(\mathcal{Z}, D)$ forms a metric space in the sense that D is a well-defined distance measure on $\mathcal{Z}$ (Lemma 4, Bathia et al. (2010)).

(vii) For any $L \in \mathcal{S}$, note that

$$\|L\|_{\mathcal{S}} = \sqrt{\mathrm{tr(L^* L)}},$$

where tr denotes the trace operator. Now, for any $\chi_i \in \mathcal{Z}$ $(i = 1, 2, 3)$, let Π_{χ_i} denote its corresponding d_0-dimensional projection operators defined as follows

$$\Pi_{\chi_i} = \sum_{j=1}^{d_0} \zeta_{ij} \otimes \zeta_{ij} \tag{7.2}$$

where $\{\zeta_{ij} : j = 1, \ldots, d_0\}$ is some orthonormal basis of χ_i.

7.2. Proof of Lemma 2.1

For the sake of convenience, let

$$\mathcal{V}_{d,t}(u) = \sum_{j=1}^{d} \eta_{tj} \psi_j(u) \quad \text{and} \quad \mathcal{V}_t(u) = \sum_{j=1}^{\infty} \eta_{tj} \psi_j(u).$$

Let us begin by noting that $E[\mathcal{V}_{d,t}(u)\mathcal{V}_{d,t+q}(v)]$ reduces to $E[\mathcal{V}_{d,t}(u)\mathcal{V}_{d,t}(v)] \equiv E[\vartheta_{d,t}(u)\vartheta_{d,t}(v)]$ when $q = 0$. Similarly,

$$\mathrm{M^{(q)}} = \sum_{i,j=1}^{d} \sigma_{ij}^{(q)} \varphi_i \otimes \varphi_j = \sum_{i=1}^{d} \lambda_i^{(q)} \varphi_i \otimes \rho_i^{(q)},$$

where $\rho_i^{(q)} = \frac{\sum_{j=1}^{d} \sigma_{ij}^{(q)} \varphi_j}{\left\|\sum_{j=1}^{d} \sigma_{ij}^{(q)} \varphi_j\right\|}$ and $\lambda_k^{(q)} = \left\|\sum_{j=1}^{d} \sigma_{ij}^{(q)} \varphi_i\right\|$, reduces to

$$\mathrm{M^{(0)}} = \sum_{i,j=1}^{d} \sigma_{ij}^{(0)} \varphi_i \otimes \varphi_j = \sum_{i=1}^{d} \lambda_i \varphi_i \otimes \varphi_i.$$

Now observe that

$$E|\mathcal{V}_{d,t}(u) - \mathcal{V}_t(u)|^2 = E[\mathcal{V}_{d,t}^2(u)] - 2E[\mathcal{V}_{d,t}(u)\mathcal{V}_t(u)] + E[\mathcal{V}_t^2(u)].$$

In this regard, the above arguments suggest that

$$\begin{aligned} E[\mathcal{V}_{d,t}^2(u)] &= E\left[\left(\sum_{i=1}^{d} \xi_{ti}\varphi_i(u)\right)\left(\sum_{j=1}^{d} \xi_{tj}\varphi_j(u)\right)\right] = \sum_{i,j=1}^{d} \varphi_i(u)\varphi_j(u)E[\xi_{ti}\xi_{tj}] \\ &= \sum_{k=1}^{d} \lambda_k \varphi_k^2(u) \end{aligned}$$

and

$$E[\mathcal{V}_{d,t}(u)\mathcal{V}_t(u)] = E\left[\left(\sum_{j=1}^{d} \xi_{tj}\varphi_j(u)\right)\mathcal{V}_t(u)\right] = \sum_{j=1}^{d} \varphi_j(u)E[\xi_{tj}\mathcal{V}_t(u)].$$

Accordingly,

$$E|\mathcal{V}_{d,t}(u) - \mathcal{V}_t(u)|^2 = \sum_{k=1}^{d} \lambda_k \varphi_k^2(u) - 2\sum_{j=1}^{d} \varphi_j(u)E[\xi_{tj}\mathcal{V}_t(u)] + E[\mathcal{V}_t^2(u)].$$

With regard to the second term, observe that

$$E[\xi_{tj}\mathcal{V}_t(u)] = E\left[\mathcal{V}_t(u)\int_{\mathcal{I}}\mathcal{V}_t(v)\varphi_j(v)dv\right] = \int_{\mathcal{I}} M^{(0)}(u,v)\varphi_j(v)dv \;=\; \lambda_j\varphi_j(u). \quad (7.3)$$

As the results,

$$\begin{aligned} E|\mathcal{V}_{d,t}(u)-\mathcal{V}_t(u)|^2 &= E[\mathcal{V}_t^2(u)] + \sum_{j=1}^d \lambda_j\varphi_j^2(u) - 2\sum_{j=1}^d \lambda_j\varphi_j^2(u) \\ &= E[\mathcal{V}_t^2(u)] - \sum_{j=1}^d \lambda_j\varphi_j^2(u) \underset{d\to\infty}{\longrightarrow} 0. \quad (7.4) \end{aligned}$$

uniformly in $u\in\mathcal{I}$. Such a convergence follows directly from the Mercer's Theorem. (See e.g. Appendix 7.1, Mercer (1909), Porter and Stirling (1990), for details.)

7.3. Proof of Theorem 2.1

Let us begin with a list of assumptions. These are standard and can be found in studies on the kernel estimation of dependence data; see, for example, Fan and Yao (2003), and Hansen (2008).

Assumption 7.1. *(a) Let $f_{U,t}(\cdot)$ and $f_{s,t}(\cdot,\cdot)$ denote the marginal density of $U_{t,\tau}$ and joint density of $(U_{t,\tau},U_{t,\tau+s})$, respectively. Assume that $f_{U,t}(\cdot)$ has a bounded support, e.g. $[c,d]$. In addition: (i) $f_{U,t}(u)>0$, $|f_{U,t}(u)-f_{U,t}(u')|\le\Delta_1|u-u'|$ for $u,u'\in[c,d]$ and some $\Delta_1>0$; (ii) $f_{s,t}(u_0,u_s)>0$ for $u_0,u_s\in[c,d]$; (iii) $\sup_{u\in[c,d]} f_{U,t}(u)\le L_0<\infty$ and $\sup_{u_0,u_s\in[c,d]} f_{s,t}(u_0,u_s)\le L_1<\infty$.*

(b) For $t=1,\dots,n$, $\{(r_{k,t,\tau},r_{k,t,\tau},U_{t,\tau}):\tau=1,\dots,m\}$ are strictly stationary and strong mixing time series with coefficient $\alpha(N)\le CN^{-\beta}$ for some $C>0$, $\beta>2+\frac{2}{\delta}$ and $\delta>0$. In addition: $E|r_{k,t,\tau}|^{4(1+\delta)}\le L_2<\infty$ and $E|r_{\ell,t,\tau}|^{4(1+\delta)}\le L_2<\infty$.

(c) Assume that $\mu_{k\ell,t}(u)$, $\mu_{k,t}(u)$, $\mu_{\ell,t}(u)$, $\sigma^2_{k,t}(u)$ and $\sigma^2_{\ell,t}(u)$ are differentiable, while $\mu''_{k\ell,t}(u)$, $\mu''_{k,t}(u)$, $\mu''_{\ell,t}(u)$, $\sigma^{2''}_{k,t}(u)$ and $\sigma^{2''}_{\ell,t}(u)$ are uniformly continuous.

(d) Assume that $\kappa(\cdot)$ is continuous symmetric kernel function, while $\int|\kappa(v)|dv<\infty$, $\int\kappa^2(v)dv<\infty$, $\int\kappa(v)dv=1$, $\int v\kappa(v)dv=0$, $\int v^2\kappa(v)dv=w_2^\kappa$ and $\int\kappa^2(v)dv=\nu_\kappa^2$. For some $0<C_1<\infty$ and $0<\Delta_2<\infty$, either $\kappa(\cdot)$ is a bounded function with a bounded support on $\mathbb{R}$ (such as $[-C_1,C_1]$), satisfying the Lipschitz condition, $|\kappa(v_1)-\kappa(v_2)|\le\Delta_2|v_1-v_2|$, or $\kappa(\cdot)$ is differentiable, when $v\to\infty$, $\kappa(v)e^{c_0v}\to 0$ $(c_0>0)$.

(e) Suppose $\frac{m}{h^2}\left(\frac{\log m}{mh}\right)^{\frac{\beta\delta-1}{2(\delta+1)}}=o(1)$ and $h=\{\log m/m\}^{1/5}$, which is allowed for sufficiently large β.

Lemma 7.1 below present uniform convergence rates that will be useful for the proof that follows.

Lemma 7.1. *Under the conditions of Assumption 7.1 and $r_{k,t,\tau}=\mu_{k,t}(U_{t,\tau})+\sigma_{k,t}(U_{t,\tau})\epsilon_{k,t,\tau}$ for $\tau=1,\dots,m$, where $E\{\epsilon_{k,t,\tau}|U_{t,\tau}\}=0$. In addition, let $\hat\mu_{k,t}(u)$ denote the local linear estimator of $\mu_{k,t}(u)$. Then:*

(i) *We have uniformly*

$$\hat{\mu}_{k,t}(u) = \mu_{k,t}(u) + \frac{1}{2} w_2^{\kappa} \mu''_{k,t}(u) h^2 + N_1(u) + \delta_m, \tag{7.5}$$

where $N_1(u) = \frac{1}{m f_{U,t}(u)} \sum_{s=1}^{m} \kappa_h(U_{t,s} - u)\sigma_{k,t}(U_{t,s})\epsilon_{k,t,\tau}$ *and* $\delta_m = o_P(h^2 + \{\log m/(mh)\}^{1/2})$.

(ii) *In addition:*

$$\sup_{u\in[c,d]} |A_1(u)| = O_p(\{\log m/(mh)\}^{1/2}), \sup_{u,v\in[c,d]} |A_2(u)| = O_P(\frac{1}{h}_t \{\log m/(mh)\}^{1/2}),$$

where $A_1(u) = \frac{1}{m} \sum_{s=1}^{m} \left[\kappa_h(U_{t,s} - u) r_{k,t,s} - E\left\{\kappa_h(U_{t,s} - u) r_{k,t,s}\right\}\right]$ *and* $A_2(u) = \frac{1}{m} \sum_{s=1}^{m} [\kappa_h(U_{t,s} - u)\kappa_h(U_{t,s} - v) r_{k,t,s} - E\left\{\kappa_h(U_{t,s} - u)\kappa_h(U_{t,s} - v) r_{k,t,s}\right\}]$.

These results are well-known and their proof can be found in studies on the uniform convergence properties for kernel estimation with dependent data (see, for example, Fan (1996), Fan and Yao (2003) and Hansen (2008)).

Similar uniform convergence rates can be obtained for those local linear estimators that are involved in $\hat{\rho}_t(u)$ in (2.24). For convenience, let $\kappa_{h,t,\tau}(u) \equiv \kappa_h(U_{t,\tau} - u)$.

(a) Regarding the local linear estimator of $\mu_{k,t}(u)\mu_{\ell,t}(u)$, it is the case that

$$\begin{aligned} &\hat{\mu}_{k,t}(u)\hat{\mu}_{\ell,t}(u) - \mu_{k,t}(u)\mu_{\ell,t}(u) \\ &\quad = \frac{1}{2} w_2^{\kappa} \{\mu_{k,t}(u)\mu''_{\ell,t}(u) + \mu''_{k,t}(u)\mu_{\ell,t}(u)\} h^2 + N_2(u) + \delta_m \end{aligned} \tag{7.6}$$

uniformly, where

$$N_2(u) \;=\; \frac{1}{m f_{U,t}(u)} \sum_{s=1}^{m} \kappa_{h,t,\tau}(u) \{\mu_{\ell,t}(u)\sigma_{k,t}(U_{t,s})\epsilon_{k,t,s} + \mu_{k,t}(u)\sigma_{\ell,t}(U_{t,s})\epsilon_{\ell,t,s}\}.$$

(b) Regarding the local linear estimator of $\sigma^2_{k,t}(u)$, we have

$$\hat{\sigma}^2_{k,t}(u) = \sigma^2_{k,t}(u) + \frac{1}{2} w_2^{\kappa} \sigma^{2''}_{k,t}(u) h^2 + N_3(u) + \delta_m \tag{7.7}$$

uniformly, where

$$N_3(u) = \frac{1}{m f_{U,t}(u)} \sum_{s=1}^{m} \kappa_{h,t,\tau}(u) \sigma^2_{k,t}(U_{t,s}) \xi_{k,t,s}$$

and $\xi_{k,t,s} = \epsilon^2_{k,t,s} - 1$. In addition, we can also obtain based on (7.7)

$$\begin{aligned} \frac{1}{\sqrt{\hat{\sigma}^2_{k,t}(u)\hat{\sigma}^2_{\ell,t}(u)}} \;=\; & \frac{1}{\sqrt{\sigma^2_{k,t}(u)\sigma^2_{\ell,t}(u)}} \Bigg[1 - w_2^{\kappa} \left(\frac{(\sigma^2_{k,t}(u))''}{4\sigma^2_{k,t}(u)} + \frac{(\sigma^2_{\ell,t}(u))''}{4\sigma^2_{\ell,t}(u)} \right) h^2 \\ & - \frac{1}{m f_U(u)} \sum_{s=1}^{m} \kappa_{h,t,\tau}(u) \left(\frac{\sigma^2_{k,t}(U_{t,s})\xi_{k,t,s}}{2\sigma^2_{k,t}(u)} + \frac{\sigma^2_{\ell,t}(U_{t,s})\xi_{\ell,t,s}}{2\sigma^2_{\ell,t}(u)} \right) \Bigg] + \delta_m. \end{aligned} \tag{7.8}$$

(c) Regarding the local linear estimator of $\mu_{k\ell,t}(u)$, we have

$$\hat{\mu}_{k\ell,t}(u) = \mu_{k\ell,t}(u) + \frac{1}{2} w_2^{\kappa} \mu''_{k\ell,t}(u) h^2 + N_4(u) + \delta_m \tag{7.9}$$

uniformly, where

$$N_4(u) = \frac{1}{mf_{U,t}(u)} \sum_{\tau=1}^{m} \kappa_{h,t,\tau}(u)\tilde{e}_{k\ell,t,\tau}$$

and $\tilde{e}_{k\ell,t,\tau} = r_{\ell,t,\tau} r_{k,t,\tau} - E(r_{\ell,t,\tau} r_{k,t,\tau} | U_{t,\tau} = u)$.

(d) Regarding the local linear estimator of $\mu_{k\ell,t}(u) - \mu_{k,t}(u)\mu_{\ell,t}(u)$, we have

$$\begin{aligned}\hat{\mu}_{k\ell,t}(u) - \hat{\mu}_{\ell,t}(u)\hat{\mu}_{k,t}(u) &= \mu_{k\ell,t}(u) - \mu_{\ell,t}(u)\mu_{k,t}(u) \\ &+ \frac{1}{2} w_2^K \big[\mu''_{k\ell,t}(u) - \mu_{k,t}(u)\mu''_{\ell,t}(u) - \mu_{\ell,t}(u)\mu''_{k,t}(u)\big] h^2 + N_5(u) + \delta_m,\end{aligned} \tag{7.10}$$

uniformly, where

$$N_5(u) = \frac{1}{mf_{U,t}(u)} \sum_{s=1}^{m} \kappa_{h,t,\tau}(u) e_{k\ell,t,s}$$

and

$$\begin{aligned}e_{k\ell,t,s} = (r_{k,t,s} - \mu_{k,t}(U_{t,s}))(r_{\ell,t,s} - \mu_{\ell,t}(U_{t,s})) \\ - E\{(r_{k,t,s} - \mu_{k,t}(U_{t,s}))(r_{\ell,t,s} - \mu_{\ell,t}(U_{t,s})) | U_{t,s}\}.\end{aligned}$$

Proof of Theorem 2.1. Regarding the local linear estimator of $\rho_t(u)$, results (a) to (d) above suggest that we have

$$\hat{\rho}_t(u) = \rho_t(u) + \frac{1}{2} w_2^\kappa B_{1\hat{\rho}}(u) h^2 - \frac{1}{2} w_2^\kappa B_{2\hat{\rho}}(u) h^2 + N_{\hat{\rho}}(u) + \delta_m, \tag{7.11}$$

uniformly, where $\delta_m = o_P(h^2 + \{\log m/(mh)\}^{1/2}$,

$$B_{1\hat{\rho}}(u) = \frac{\mu''_{k\ell,t}(u) - \mu_{k,t}(u)\mu''_{\ell,t}(u) - \mu_{\ell,t}(u)\mu''_{k,t}(u)}{\sigma_{\ell,t}(u)\sigma_{k,t}(u)},$$

$$B_{2\hat{\rho}}(u) = \frac{\rho_t(u)(\sigma^2_{k,t}(u))''}{2\sigma^2_{k,t}(u)} - \frac{\rho_t(u)(\sigma^2_{\ell,t}(u))''}{2\sigma^2_{\ell,t}(u)},$$

$$N_{\hat{\rho}}(u) = \frac{1}{mf_{U,t}(u)} \sum_{s=1}^{m} \kappa_{h,t,\tau}(u) N_{\hat{\rho},\tau}(u)$$

and

$$N_{\hat{\rho},s}(u) = \frac{e_{k\ell,t,s}}{\sigma_{\ell,t}(u)\sigma_{k,t}(u)} - \frac{\rho_t(u)\sigma^2_{k,t}(U_{t,s})\xi_{k,t,s}}{2\sigma^2_{k,t}(u)} - \frac{\rho_t(u)\sigma^2_{\ell,t}(U_{t,s})\xi_{\ell,t,s}}{2\sigma^2_{\ell,t}(u)}.$$

Theorem 2.1 follows immediately from (7.11).

7.4. Proof of Theorems 2.2

Providing the proof for Theorems 2.2 requires some additional conditions as follows.

Assumption 7.2. *(i) Assumption 7.1 holds.*

(ii) The FC-TS, $\{\rho_t(\cdot)\}$, is strictly stationary and ψ-mixing with mixing coefficient defined as

$$\psi(l) = \sup_{A\in\mathcal{F}_{\infty}^{0},\ B\in\mathcal{F}_{l}^{\infty},\ P(A)P(B)>0} \left|1 - \frac{P(B|A)}{P(B)}\right|,$$

where $\mathcal{F}_i^j = \sigma\{\rho_i(\cdot), \dots \rho_j(\cdot)\}$ for any $j \geq i$ and $\sum_{l=1}^{\infty} l \times \psi^{1/2}(l) < \infty$.

(iii) The FC-TS is square integrable curve series, i.e.

$$E\left\{\int_{\mathcal{I}} \rho_t(u)^2 \ du\right\}^2 < \infty \ \textit{and} \ \int_{\mathcal{I}} E\{\vartheta_t(u)^2\} \ du < \infty.$$

(iv) All nonzero eigenvalues of K *are different.*

Moreover, the following observations will be useful at various stages of the proof.

(a) Since $N^{(q)}(u,v) = \int_{\mathcal{I}} M^{(q)}(u,z)M^{(q)}(v,z)dz$, we have

$$\begin{aligned}(N^{(q)}f)(u) &= \int N^{(q)}(u,v)f(v) \ dv \\ &= \sum_{i,j=1}^{\infty} w_{ij}^{(q)}\langle\varphi_i, f\rangle\varphi_j(u) \ = \ (\mathrm{M}^{(q)}\mathrm{M}^{(q)*}f)(u),\end{aligned}$$

which suggests therefore that $\mathrm{N}^{(q)} = \mathrm{M}^{(q)}\mathrm{M}^{(q)*}$.

(b) For convenience, let $\hat{\rho}_t(u) - \rho_t(u) = \Delta_{\hat{\rho},\rho}$. In this regard, Theorem 2.1 and the bandwidth given in Assumption 7.1(e) suggest that

$$\Delta_{\hat{\rho},\rho} = O_P((\log m/m)^{2/5}). \tag{7.12}$$

Since

$$n^{1/2} = \left\lfloor \left(\frac{m}{\log m}\right)^{2/5} \right\rfloor \tag{7.13}$$

as required in condition (2.37), $n^{1/2} \leq (m/\log m)^{2/5}$, then it must be the case that

$$\left(\frac{\log m}{m}\right)^{2/5} \leq \frac{1}{n^{1/2}}. \tag{7.14}$$

In other words,

$$\Delta_{\hat{\rho},\rho} \leq O_P(n^{-1/2}). \tag{7.15}$$

(c) With regard to the expected correlation $\varrho(u) = E\{\rho_t(u)\}$, we have considered a pair of estimators, namely

$$\tilde{\varrho}(u) = n^{-1}\sum_{1\leq j\leq n} \rho_j(u) \ \text{ and } \ \hat{\varrho}(u) = n^{-1}\sum_{1\leq j\leq n} \hat{\rho}_j(u).$$

Here, observe that

$$|\hat{\varrho}(u) - \varrho(u)| \leq |\hat{\varrho}(u) - \tilde{\varrho}(u)| + |\tilde{\varrho}(u) - \varrho(u)|,$$

where $|\tilde{\varrho}(u)-\varrho(u)| = O_P(n^{-1/2})$ following a simple U-statistic argument (see Lee (1990)). Regarding the first term, we have

$$\begin{aligned}|\hat{\varrho}(u)-\tilde{\varrho}(u)| &\leq n^{-1}\sum_{t=1}^{n}|\hat{\rho}_t(u)-\rho_t(u)| = |\Delta_{\hat{\rho},\rho}| \\ &\leq O_P(n^{-1/2}),\end{aligned}$$

where the second inequality is due to (7.14). Observe also that

$$\begin{aligned}\{\hat{\rho}_j(u)-\hat{\varrho}(u)\}\{\hat{\rho}_{j+q}(u)-\hat{\varrho}(u)\} \;&\leq\; \{\hat{\rho}_j(u)-\varrho(u)\}\{\hat{\rho}_{j+q}(v)-\varrho(v)\} \\ &+ |\varrho(u)-\hat{\varrho}(u)||\varrho(v)-\hat{\varrho}(v)| + |\varrho(u)-\hat{\varrho}(u)||\hat{\rho}_{j+q}(v)-\varrho(v)| \\ &+ |\hat{\rho}_j(u)-\varrho(u)||\varrho(v)-\hat{\varrho}(v)| \\ &= \{\hat{\rho}_j(u)-\varrho(u)\}\{\hat{\rho}_{j+q}(v)-\varrho(v)\} \\ &+ |\Delta_{\hat{\rho},\rho}|^2 + |\hat{\rho}_{j+q}(v)-\varrho(v)||\Delta_{\hat{\rho},\rho}| \;+\; |\hat{\rho}_j(u)-\varrho(u)||\Delta_{\hat{\rho},\rho}|. \qquad (7.16)\end{aligned}$$

Without loss of generality, results (7.15) and (7.16) suggest that we can consider $\{\hat{\rho}_j(u)-\varrho(u)\}\{\hat{\rho}_{j+q}(v)-\varrho(v)\}$ instead of $\{\hat{\rho}_j(u)-\hat{\varrho}(u)\}\{\hat{\rho}_{j+q}(u)-\hat{\varrho}(u)\}$ in the remaining of the proof.

(d) Furthermore:

$$\begin{aligned}\{\hat{\rho}_j(u)-\varrho(u)\}\{\hat{\rho}_{j+q}(v)-\varrho(v)\} &\leq \{\rho_j(u)-\varrho(u)\}\{\rho_{j+q}(v)-\varrho(v)\} \\ &+ |\hat{\rho}_j(u)-\rho_j(u)||\hat{\rho}_{j+q}(v)-\rho_{j+q}(v)| + |\rho_j(u)-\varrho(u)||\hat{\rho}_{j+q}(v)-\rho_{j+q}(v)| \\ &+ |\hat{\rho}_j(u)-\rho_j(u)||\rho_{j+q}(v)-\varrho(v)| \\ &= \{\rho_j(u)-\varrho(u)\}\{\rho_{j+q}(v)-\varrho(v)\} \\ &+ |\Delta_{\hat{\rho},\rho}|^2 + |\rho_j(u)-\varrho(u)||\Delta_{\hat{\rho},\rho}| + |\rho_{j+q}(v)-\varrho(v)||\Delta_{\hat{\rho},\rho}| \qquad (7.17)\end{aligned}$$

(e) Let $\tilde{Z}_{tq}(u,v) = \{\rho_t(u)-\varrho(u)\}\{\rho_{t+q}(v)-\varrho(v)\}$. In this regard,

$$\begin{aligned}\tilde{Z}_{iq}\tilde{Z}^*_{jq}(u,v) &= \int_{\mathcal{I}}\tilde{Z}_{iq}(u,r)\tilde{Z}_{jq}(v,r)\;dr \\ &= \{\rho_i(u)-\varrho(u)\}\{\rho_j(v)-\varrho(v)\}\langle\rho_{i+q}-\varrho,\rho_{j+q}-\varrho\rangle. \qquad (7.18)\end{aligned}$$

Furthermore,

$$\int_{\mathcal{I}}\tilde{Z}_{iq}\tilde{Z}^*_{jq}(u,v)f(v)\;dv = \{\rho_i(u)-\varrho(u)\}\langle\rho_j-\varrho,f\rangle\langle\rho_{i+q}-\varrho,\rho_{j+q}-\varrho\rangle. \qquad (7.19)$$

It is therefore the case that

$$\tilde{\mathrm{Z}}_{ik}\tilde{\mathrm{Z}}^*_{jk} = (\rho_i-\varrho)\otimes(\rho_j-\varrho)\langle\rho_{i+q}-\varrho,\rho_{j+q}-\varrho\rangle. \qquad (7.20)$$

Accordingly, one can write

$$\tilde{\mathrm{M}}^{(\mathrm{q})}\tilde{\mathrm{M}}^{(\mathrm{q})^*} = \frac{1}{(n-p)^2}\sum_{i,j=1}^{n-p}\tilde{\mathrm{Z}}_{ik}\tilde{\mathrm{Z}}^*_{jk}, \qquad (7.21)$$

which is a $\mathcal{S}$ valued von Mises functional. In this regard, Lemma 3 of Bathia et al. (2010) suggests that we have

$$E\|\tilde{\mathrm{M}}^{(\mathrm{q})}\tilde{\mathrm{M}}^{(\mathrm{q})^*} - \mathrm{M}^{(\mathrm{q})}\mathrm{M}^{(\mathrm{q})^*}\|^2_{\mathcal{S}} = O(n^{-1}). \qquad (7.22)$$

(f) Given the definition in (2.36), we can also construct $\hat{\mathrm{N}}^{(\mathrm{q})} = \hat{\mathrm{M}}^{(\mathrm{q})}\hat{\mathrm{M}}^{(\mathrm{q})*}$ by following a similar procedure to that in point (e). Then, this leads to

$$\hat{\mathrm{K}} \ = \ \sum_{q=1}^{p} \hat{\mathrm{M}}^{(\mathrm{q})}\hat{\mathrm{M}}^{(\mathrm{q})*}. \tag{7.23}$$

(g) Let us recall

$$\hat{\mathbf{K}}^{*}\hat{\boldsymbol{\gamma}}_j = \hat{\boldsymbol{\gamma}}_j\hat{\theta}_j$$

from just above (2.34). Decomposing this component by component leads to

$$\frac{1}{(n-p)^2}\sum_{t,s=1}^{n-p}\sum_{k=1}^{p}\langle\hat{\rho}_{t+q}-\varrho,\hat{\rho}_{s+q}-\varrho\rangle\langle\hat{\rho}_{s}-\varrho,\hat{\rho}_{t}-\varrho\rangle\hat{\gamma}_{tj} = \hat{\gamma}_{tj}\hat{\theta}_j. \tag{7.24}$$

Regarding $\langle\hat{\rho}_{t+q}-\varrho,\hat{\rho}_{s+q}-\varrho\rangle$, a similar decomposition to (7.17) together with Theorem 2.1 and the bandwidth given in Assumption 7.1(e) suggest

$$\begin{aligned}&\int(\hat{\rho}_{t+q}(u)-\varrho(u))(\hat{\rho}_{s+q}(u)-\varrho(u))\ du\\ &\qquad = \int(\rho_{t+q}(u)-\varrho(u))(\rho_{s+q}(u)-\varrho(u))\ du + \Delta_{\hat{\rho},\rho},\end{aligned} \tag{7.25}$$

which holds for all $q = 1,\ldots,p$. A similar result can also be worked out for $\langle\hat{\rho}_{s}-\varrho,\hat{\rho}_{t}-\varrho\rangle$. We then obtain by applying these results to all components of $\hat{\mathbf{K}}^{*}$

$$\hat{\mathbf{K}}^{*} = \tilde{\mathbf{K}}^{*} + \Delta_{\hat{\rho},\rho}1_{n-p}1_{n-p}', \tag{7.26}$$

where $\tilde{\mathbf{K}}^{*}$ is as defined in (2.30) and 1_{n-p} is a column vector of length $n-p$. In this sense, differentiation using the results in Magnus (1985) and the Taylor's expansion in a similar fashion to the proof of Theorem 3.5 of Jiang et al. (2016) lead to

$$\hat{\theta}_j - \tilde{\theta}_j \ = \ \tilde{\boldsymbol{\gamma}}_j'(\hat{\mathbf{K}}^{*}-\tilde{\mathbf{K}}^{*})\tilde{\boldsymbol{\gamma}}_j \tag{7.27}$$

$$\hat{\boldsymbol{\gamma}}_j - \tilde{\boldsymbol{\gamma}}_j \ = \ (\tilde{\theta}_j\mathbf{I}-\tilde{\mathbf{K}}^{*})^{+}(\hat{\mathbf{K}}^{*}-\tilde{\mathbf{K}}^{*})\tilde{\boldsymbol{\gamma}}_j, \tag{7.28}$$

where $\mathbf{I}$ is the identity matrix of size $n-p$ and $(\cdot)^{+}$ denotes the Moore-Penrose inverse.

Proof of Theorem 2.2 (i) We begin by writing

$$\hat{M}^{(q)}(u,v) = \tilde{M}^{(q)}(u,v) + \Delta_1(u,v), \tag{7.29}$$

where

$$\begin{aligned}\Delta_1(u,v) \ &= \ \frac{1}{n-p}\sum_{j=1}^{n-p}\Big(\{\hat{\rho}_j(u)-\hat{\varrho}(u)\}\{\hat{\rho}_{j+q}(v)-\hat{\varrho}(v)\}\\ &\qquad\qquad - \ \{\rho_j(u)-\varrho(u)\}\{\rho_{j+q}(v)-\varrho(v)\}\Big)\\ &\leq \ \frac{1}{n-p}\sum_{j=1}^{n-p}\Big(\{\hat{\rho}_j(u)-\varrho(u)\}\{\hat{\rho}_{j+q}(v)-\varrho(v)\} - \{\rho_j(u)-\varrho(u)\}\{\rho_{j+q}(v)-\varrho(v)\}\Big)\\ &\quad + |\Delta_{\hat{\rho},\rho}|^2 + |\hat{\rho}_{j+q}(v)-\varrho(v)||\Delta_{\hat{\rho},\rho}| \ + \ |\hat{\rho}_j(u)-\varrho(u)||\Delta_{\hat{\rho},\rho}|,\end{aligned} \tag{7.30}$$

where the inequality is due to (7.16). Accordingly, the finding in point (c) above suggests that it is reasonable to focus instead on

$$\Delta_1(u,v) = \frac{1}{n-p}\sum_{j=1}^{n-p}\Big(\{\hat\rho_j(u)-\varrho(u)\}\{\hat\rho_{j+q}(v)-\varrho(v)\}-\{\rho_j(u)-\varrho(u)\}\{\rho_{j+q}(v)-\varrho(v)\}\Big).$$

This can be written as $\Delta_1(u,v) = \Delta_{11}(u,v) + \Delta_{12}(u,v) + \Delta_{13}(u,v)$ in which

$$\begin{aligned}
\Delta_{11}(u,v) &= \frac{1}{n-p}\sum_{j=1}^{n-p}\{\hat\rho_j(u)-\rho_j(u)\}\{\hat\rho_{j+q}(v)-\rho_{j+q}(v)\}\\
\Delta_{12}(u,v) &= \frac{1}{n-p}\sum_{j=1}^{n-p}\{\rho_j(u)-\varrho(u)\}\{\hat\rho_{j+q}(v)-\rho_{j+q}(v)\}\\
\Delta_{13}(u,v) &= \frac{1}{n-p}\sum_{j=1}^{n-p}\{\hat\rho_j(u)-\rho_j(u)\}\{\rho_{j+q}(v)-\varrho(v)\}.
\end{aligned}$$

Such a decomposition leads to

$$\hat K(u,v) = \sum_{q=1}^{p}\int \hat M^{(q)}(u,z)\hat M^{(q)}(v,z)\,dz = \tilde K(u,v) + \Delta_2(u,v), \tag{7.31}$$

where

$$\tilde K(u,v) = \sum_{q=1}^{p}\int \tilde M^{(q)}(u,z)\tilde M^{(q)}(v,z)\,dz$$

and

$$\begin{aligned}
\Delta_2(u,v) &= \sum_{q=1}^{p}\int \Delta(u,z)\Delta(v,z)dz + \sum_{q=1}^{p}\int \Delta(u,z)\tilde M^{(q)}(v,z)\,dz\\
&+ \sum_{q=1}^{p}\int \tilde M^{(q)}(u,z)\Delta(v,z)\,dz.
\end{aligned} \tag{7.32}$$

Then, for $\hat\psi_j$ computed based on (2.34), we write

$$\int_{\mathcal I}\hat K(u,v)\hat\psi_j(v)\,dv = \int_{\mathcal I}\{\tilde K(u,v)+\Delta_2(u,v)\}\hat\psi_j(v)\,dv. \tag{7.33}$$

Moreover, since

$$\begin{aligned}
\hat\gamma_{tj}\{\hat\rho_t(u)-\varrho(u)\} &= \hat\gamma_{tj}\{[\hat\rho_t(u)-\rho_t(u)]+[\rho_t(u)-\varrho(u)]\}\\
&= (\tilde\gamma_{tj}+\Delta_{\hat\rho,\rho})\{[\hat\rho_t(u)-\rho_t(u)]+[\rho_t(u)-\varrho(u)]\}\\
&= \tilde\gamma_{tj}\{\rho_t(u)-\varrho(u)\}+\Delta_{\hat\rho,\rho}
\end{aligned}$$

under (7.28), the first term of (7.33) is

$$\int_{\mathcal I}\tilde K(u,v)\hat\psi_j(v)\,dv = \int_{\mathcal I}\tilde K(u,v)(\tilde\psi_j(v)+\Delta_{\hat\rho,\rho})\,dv + \Delta_{\hat\rho,\rho}, \tag{7.34}$$

where $\tilde{\psi}_j(v)$ is defined in (2.31). To break $\int_{\mathcal{I}} \Delta_2(u,v)\hat{\psi}_j(v)\ dv$ down, let us consider the third term on the right side of (7.32) as an example. In this respect,

$$\begin{aligned}
&\int_{\mathcal{I}} \tilde{M}^{(q)}(u,z)\Delta_1(v,z)\ dz \\
&\qquad = \frac{1}{(n-p)^2}\sum_{t,s=1}^{n-p}\{\rho_t(u)-\varrho(u)\}\{\hat{\rho}_j(v)-\rho_j(v)\}\langle \rho_{t+q}-\varrho, \hat{\rho}_{s+q}-\rho_{s+q}\rangle \\
&\int_{\mathcal{I}} \tilde{M}^{(q)}(u,z)\Delta_2(v,z)\ dz \\
&\qquad = \frac{1}{(n-p)^2}\sum_{t,s=1}^{n-p}\{\rho_t(u)-\varrho(u)\}\{\hat{\rho}_j(v)-\rho_j(v)\}\langle \rho_{t+q}-\varrho, \hat{\rho}_{s+q}-\rho_{s+q}\rangle \\
&\int_{\mathcal{I}} \tilde{M}^{(q)}(u,z)\Delta_3(v,z)\ dz \\
&\qquad = \frac{1}{(n-p)^2}\sum_{t,s=1}^{n-p}\{\rho_t(u)-\varrho(u)\}\{\hat{\rho}_j(v)-\rho_j(v)\}\langle \rho_{t+q}-\varrho, \rho_{s+q}-\varrho\rangle.
\end{aligned}$$

Hence, Theorem 2.1 and the bandwidth given in Assumption 7.1(e) suggest that

$$\int \tilde{M}^{(q)}(u,z)\Delta(v,z)\ dz = \Delta_{\hat{\rho},\rho}, \tag{7.35}$$

which holds for all $q = 1,\dots,p$. The rest of the terms can be similarly worked out. These results suggest that

$$\int_{\mathcal{I}} \hat{K}(u,v)\hat{\psi}_j(v)\ dv \ = \ \int_{\mathcal{I}} \tilde{K}(u,v)\tilde{\psi}_j(v)\ dv + \Delta_{\hat{\rho},\rho}. \tag{7.36}$$

By making use of (7.23) and taking into consideration the definition in (2.23), we write

$$\hat{\mathrm{K}} \ = \ \sum_{q=1}^{p}\hat{\mathrm{M}}^{(\mathrm{q})}\hat{\mathrm{M}}^{(\mathrm{q})^*} \ = \ \sum_{q=1}^{p}\left(\tilde{\mathrm{M}}^{(\mathrm{q})}\tilde{\mathrm{M}}^{(\mathrm{q})^*} + (1/p)\Delta_{\hat{\rho},\rho}\right), \tag{7.37}$$

where $\tilde{\mathrm{N}}^{(\mathrm{q})} = \tilde{\mathrm{M}}^{(\mathrm{q})}\tilde{\mathrm{M}}^{(\mathrm{q})^*}$. In other words, we have for a given q

$$\hat{\mathrm{M}}^{(\mathrm{q})}\hat{\mathrm{M}}^{(\mathrm{q})^*} - \tilde{\mathrm{M}}^{(\mathrm{q})}\tilde{\mathrm{M}}^{(\mathrm{q})^*} = (1/p)\Delta_{\hat{\rho},\rho}. \tag{7.38}$$

Moreover, since

$$\{\hat{\mathrm{M}}^{(\mathrm{q})}\hat{\mathrm{M}}^{(\mathrm{q})^*} - \mathrm{M}^{(\mathrm{q})}\mathrm{M}^{(\mathrm{q})^*}\} = \{\tilde{\mathrm{M}}^{(\mathrm{q})}\tilde{\mathrm{M}}^{(\mathrm{q})^*} - \mathrm{M}^{(\mathrm{q})}\mathrm{M}^{(\mathrm{q})^*}\} + \{\hat{\mathrm{M}}^{(\mathrm{q})}\hat{\mathrm{M}}^{(\mathrm{q})^*} - \tilde{\mathrm{M}}^{(\mathrm{q})}\tilde{\mathrm{M}}^{(\mathrm{q})^*}\},$$

the Triangle inequality suggests that

$$\begin{aligned}
E\|\hat{\mathrm{M}}^{(\mathrm{q})}\hat{\mathrm{M}}^{(\mathrm{q})^*} - \mathrm{M}^{(\mathrm{q})}\mathrm{M}^{(\mathrm{q})^*}\|_{\mathcal{S}}^2 \quad &\leq \quad E\|\tilde{\mathrm{M}}^{(\mathrm{q})}\tilde{\mathrm{M}}^{(\mathrm{q})^*} - \mathrm{M}^{(\mathrm{q})}\mathrm{M}^{(\mathrm{q})^*}\|_{\mathcal{S}}^2 \\
&\qquad + E\|\hat{\mathrm{M}}^{(\mathrm{q})}\hat{\mathrm{M}}^{(\mathrm{q})^*} - \tilde{\mathrm{M}}^{(\mathrm{q})}\tilde{\mathrm{M}}^{(\mathrm{q})^*}\|_{\mathcal{S}}^2.
\end{aligned} \tag{7.39}$$

Regarding the first term, (7.22) and the Chebyshev inequality lead to

$$\|\tilde{\mathrm{M}}^{(\mathrm{q})}\tilde{\mathrm{M}}^{(\mathrm{q})^*} - \mathrm{M}^{(\mathrm{q})}\mathrm{M}^{(\mathrm{q})^*}\|_{\mathcal{S}} \leq O_P(n^{-1/2}). \tag{7.40}$$

Given the result in (7.38), the second term can be viewed as a compact linear operator

$$E\|\hat{\mathrm{M}}^{(\mathrm{q})}\hat{\mathrm{M}}^{(\mathrm{q})^*} - \tilde{\mathrm{M}}^{(\mathrm{q})}\tilde{\mathrm{M}}^{(\mathrm{q})^*}\|_{\mathcal{S}}^2 \leq O_P(n^{-1}), \tag{7.41}$$

where the inequality is due to (7.15). A similar application of the Chebyshev inequality to (7.40) also gives

$$\|\hat{\mathrm{M}}^{(\mathrm{q})}\hat{\mathrm{M}}^{(\mathrm{q})^*} - \tilde{\mathrm{M}}^{(\mathrm{q})}\tilde{\mathrm{M}}^{(\mathrm{q})^*}\|_{\mathcal{S}} \leq O_P(n^{-1/2}). \tag{7.42}$$

Then, the required result is obtained by writing

$$\|\hat{K} - K\|_{\mathcal{S}} \quad \leq \quad \|\hat{\mathrm{K}} - \tilde{\mathrm{K}}\|_{\mathcal{S}} + \|\tilde{K} - K\|_{\mathcal{S}} \tag{7.43}$$

and noting that

$$\|\tilde{\mathrm{K}} - \mathrm{K}\|_{\mathcal{S}} \leq O_p(n^{-1/2}) \ \text{ and } \ \|\hat{\mathrm{K}} - \tilde{\mathrm{K}}\|_{\mathcal{S}} \leq O_p(n^{-1/2}), \tag{7.44}$$

which are based on (7.40) and (7.42), respectively.

Proof of Theorems 2.2 (ii) and 2.2 (iii) The proof of there results relies on the results in (7.43) and (7.44). While $\|\tilde{\mathrm{K}} - K\|_{\mathcal{S}} = O_p(n^{-1/2})$, Lemmas 4.2 and 4.3 of Bosq (2000) suggest that

$$\sup_{j\geq 1} |\tilde{\theta}_j - \theta_j| \leq \|\tilde{\mathrm{K}} - \mathrm{K}\|_{\mathcal{S}} \text{ and } \sup_{j\geq 1} |\tilde{\psi}_j - \psi_j| \leq \|\tilde{\mathrm{K}} - \mathrm{K}\|_{\mathcal{S}}, \tag{7.45}$$

respectively. Then, Theorem 2.2 (ii) is obtained by noting (7.27) and the fact that $\|\hat{\mathrm{K}} - \tilde{\mathrm{K}}\|_{\mathcal{S}} \leq O_p(n^{-1/2})$. Given that all the nonzero eigenvalues of K are different, which is assumed in Assumption 7.2(iv), Theorem 2.2 (iii) is obtained by noting the definition in (2.34), the result in (7.28) and that $\|\hat{\mathrm{K}} - \tilde{\mathrm{K}}\|_{\mathcal{S}} \leq O_p(n^{-1/2})$.

7.5. Proof of Lemma 3.1

For the sake of convenience, let

$$e_{d_0,t}(u) = \sum_{j=d_0+1}^{\infty} \eta_{tj}\psi_j(u).$$

Observe that $E[e_{d_0,t}(u)e_{d_0,t+q}(v)]$ reduces to $E[e_{d_0,t}(u)e_{d_0,t}(v)] \equiv E[\epsilon_{d_0,t}(u)\epsilon_{d_0,t}(v)]$ when $q = 0$, where

$$\epsilon_{d_0,t}(u) = \sum_{j=d_0+1}^{\infty} \xi_{tj}\varphi_j(u).$$

These arguments suggest that the mean squared error is

$$E[e_{d_0,t}^2(u)] = \sum_{i\geq d_0+1}\sum_{j\geq d_0+1} \varphi_i(u)\varphi_j(u) \int_{\mathcal{I}}\int_{\mathcal{I}} E[\vartheta_t(t_1)\vartheta_t(s_1)]\varphi_i(t_1)\varphi_j(s_1)dt_1ds_1.$$

Integrating both sides of the equation and applying the orthogonality lead to

$$\begin{aligned}
&\int_{\mathcal{I}} E[e_{d_0,t}^2(u)]du \\
&\quad = \sum_{i\geq d_0+1}\sum_{j\geq d_0+1} \int_{\mathcal{I}} \varphi_i(u)\varphi_j(u)du \int_{\mathcal{I}}\int_{\mathcal{I}} E[\vartheta_t(t_1)\vartheta_t(s_1)]\varphi_i(t_1)\varphi_j(s_1)dt_1ds_1 \\
&\quad = \sum_{j\geq d_0+1} \int_{\mathcal{I}}\int_{\mathcal{I}} E[\vartheta_t(t_1)\vartheta_t(s_1)]\varphi_j(t_1)\varphi_j(s_1)dt_1ds_1
\end{aligned}$$

Minimising the integrated mean squared error subject to the orthogonality condition for the function of the eigenfunction, i.e.

$$\min \int_{\mathcal{I}} E[e^2_{d_0,t}(u)]du \ \text{ subject to } \int_{\mathcal{I}} \varphi_j(u)\varphi_j(u) = 1,$$

leads to the objective function

$$Q = \sum_{j \geq d_0+1} \left\{ \int_{\mathcal{I}} \int_{\mathcal{I}} M^{(0)}(t_1, s_1)\varphi_j(t_1)\varphi_j(s_1)dt_1 ds_1 - \delta_j \left(\int_{\mathcal{I}} \varphi_j(t_1)\varphi_j(t_1) - 1 \right) \right\}.$$

Differentiating Q with respect to $\varphi_i(u)$ (for $i \geq d_0 + 1$) leads to

$$\frac{d}{d\varphi_i(u)} Q = 2 \int_{\mathcal{I}} M^{(0)}(u, v)\varphi_i(v)dv - 2\lambda_i \varphi_i(u). \tag{7.46}$$

Hence, setting the above equation to zero leads to

$$(\mathrm{M}^{(0)}\varphi_i)(u) = \lambda_i \varphi_i(u), \tag{7.47}$$

which is the Fredholm integral equation. Proposition 1(ii) of Bathia et al. (2010) suggests that

$$\mathcal{V}_{d_0,t} = \sum_{j=1}^{d_0} \eta_{tj}\psi_j(u) = \sum_{j=1}^{d_0} \xi_{tj}\varphi_j(u). \tag{7.48}$$

The expansion in (7.48) has a one-to-one relationship with (7.47) and therefore minimises the integrated mean squared error.

7.6. Proof of Theorem 3.1

From Definitions (v) to (vii) given in Appendix 7.1, we have by applying the triangle inequality

$$\begin{aligned} \sqrt{2d_0} D(\hat{\mathcal{M}}, \mathcal{M}) &= \|\Pi_{\widehat{\mathcal{M}}} - \Pi_{\mathcal{M}}\|_{\mathcal{S}} \\ &\leq \|\Pi_{\widehat{\mathcal{M}}} - \Pi_{\widetilde{\mathcal{M}}}\|_{\mathcal{S}} + \|\Pi_{\widetilde{\mathcal{M}}} - \Pi_{\mathcal{M}}\|_{\mathcal{S}}, \end{aligned} \tag{7.49}$$

where $\Pi_{\widehat{\mathcal{M}}} = \sum_{j=1}^{d_0} \hat{\psi}_j \otimes \hat{\psi}_j$ and $\Pi_{\widetilde{\mathcal{M}}} = \sum_{j=1}^{d_0} \tilde{\psi}_j \otimes \tilde{\psi}_j$. Regarding the first term on the right side of the inequality, we have

$$\begin{aligned} \|\Pi_{\widehat{\mathcal{M}}} - \Pi_{\widetilde{\mathcal{M}}}\|_{\mathcal{S}} &= \|\sum_{j=1}^{d_0} \hat{\psi}_j \otimes \hat{\psi}_j - \sum_{j=1}^{d_0} \tilde{\psi}_j \otimes \tilde{\psi}_j\|_{\mathcal{S}} \ \leq \ \sum_{j=1}^{d_0} \|\hat{\psi}_j \otimes \hat{\psi}_j - \tilde{\psi}_j \otimes \tilde{\psi}_j\|_{\mathcal{S}} \\ &= O_P(n^{-1/2}) \end{aligned} \tag{7.50}$$

since $\|\hat{\mathrm{K}} - \tilde{\mathrm{K}}\|_{\mathcal{S}} \leq O_p(n^{-1/2})$. In addition,

$$\|\Pi_{\widetilde{\mathcal{M}}} - \Pi_{\mathcal{M}}\|_{\mathcal{S}} = O_P(n^{-1/2}) \tag{7.51}$$

since $\|\tilde{\psi}_j \otimes \tilde{\psi}_j - \psi_j \otimes \psi_j\|_{\mathcal{S}} = O_P(n^{-1/2})$, where the convergence rate is based on the second part of (7.45). The proof is therefore completed.

7.7. *Proof of Lemma 3.2*

Note that

$$|\hat{\mathcal{V}}_{d_0,t}(u) - \mathcal{V}_t(u)| \leq |\hat{\mathcal{V}}_{d_0,t}(u) - \mathcal{V}_{d_0,t}(u)| + |\mathcal{V}_{d_0,t}(u) - \mathcal{V}_t(u)|.$$

Lemma 2.1 implies that $\mathcal{V}_{d_0,t}(u) \overset{P}{\to} \mathcal{V}_t(u)$ as $d_0 \to \infty$. For a fixed d_0, observe that $\hat{\eta}_{tj} \overset{P}{\to} \eta_{tj}$ as $n \to \infty$, then, by Theorem 2.2(iii), $\sup_{u \in \mathcal{I}} |\hat{\mathcal{V}}_{d_0,t}(u) - \mathcal{V}_{d_0,t}(u)| \overset{P}{\to} 0$ as $n \to \infty$. For a given $\epsilon, \delta > 0$, this implies that there exists $\bar{d}$ such that for $d_0 \geq \bar{d}$,

$$P\{|\mathcal{V}_{d_0,t}(u) - \mathcal{V}_t(u)| > \epsilon/2\} \leq \delta/2.$$

For each d_0, there exists $\bar{n}(d_0)$ such that, for $n \geq \bar{n}(d_0)$,

$$P\{|\hat{\mathcal{V}}_{d_0,t}(u) - \mathcal{V}_{d_0,t}(u)| \geq \epsilon/2\} \leq \delta/2.$$

Thus, for $d_0 \geq \bar{d}$ and $n \geq \bar{n}(d_0)$

$$P\{|\hat{\mathcal{V}}_{d_0,t}(u) - \mathcal{V}_t(u)| \geq \epsilon\} \leq P\{|\hat{\mathcal{V}}_{d_0,t}(u) - \mathcal{V}_{d_0,t}(u)| \geq \epsilon/2\} + P\{|\mathcal{V}_{d_0,t}(u) - \mathcal{V}_t(u)| > \epsilon/2\} \leq \delta,$$

which leads to (3.6).

7.8. *Proof of Lemma 3.3*

Observe that

$$\begin{aligned} \hat{\theta}_j - \theta_j &= \langle \psi_j, \hat{\mathrm{K}}\hat{\psi}_j \rangle - \langle \psi_j, \mathrm{K}\psi_j \rangle \\ &= \langle \psi_j, (\hat{K} - K)\psi_j \rangle + \langle \psi_j, \hat{\mathrm{K}}\hat{\psi}_j \rangle - \langle \psi_j, \hat{\mathrm{K}}\psi_j \rangle \end{aligned} \tag{7.52}$$

We shall begin by showing that

$$\hat{\theta}_j - \theta_j = \langle \psi_j, (\hat{K} - K)\psi_j \rangle + O_P(n^{-1}) \tag{7.53}$$

for $j = 1, \ldots, d_0$.

From the second equality in (7.52),

$$\langle \psi_j, \hat{\mathrm{K}}\hat{\psi}_j \rangle - \langle \psi_j, \hat{\mathrm{K}}\psi_j \rangle = \langle \psi_j, \hat{\mathrm{K}}\hat{\psi}_j \rangle - \langle \psi_j, \mathrm{K}\hat{\psi}_j \rangle + \langle \psi_j, \mathrm{K}\hat{\psi}_j \rangle - \langle \psi_j, \hat{\mathrm{K}}\psi_j \rangle$$

by which

$$\langle \psi_j, \hat{\mathrm{K}}\hat{\psi}_j \rangle - \langle \psi_j, \mathrm{K}\hat{\psi}_j \rangle = \langle \psi_j, (\hat{K} - K)\hat{\psi}_j \rangle, \tag{7.54}$$

$$\begin{aligned} \langle \psi_j, \mathrm{K}\hat{\psi}_j \rangle - \langle \psi_j, \hat{\mathrm{K}}\psi_j \rangle &= \langle \psi_j, \mathrm{K}\hat{\psi}_j \rangle - \langle \psi_j, \mathrm{K}\psi_j \rangle + \langle \psi_j, \mathrm{K}\psi_j \rangle - \langle \psi_j, \hat{\mathrm{K}}\psi_j \rangle \\ &= \langle \psi_j, (\hat{K} - K)\hat{\psi}_j \rangle + \langle \psi_j, \mathrm{K}(\hat{\psi}_j - \psi_j) \rangle \end{aligned} \tag{7.55}$$

Let $K_j = \langle \psi_j, (\hat{K} - K)\hat{\psi}_j \rangle$ for the sake of convenience. Regarding the first term in (7.55), we want to show that, for $j = 1, \ldots, d_0$,

$$|\langle \psi_j, (\hat{K} - K)\psi_j \rangle - K_j| = O_P(n^{-1}) \tag{7.56}$$

Observe that

$$\begin{aligned} |\langle \psi_j, (\hat{K} - K)\psi_j \rangle - K_j| &= |\langle \psi_j - \hat{\psi}_j, (\hat{K} - K)\psi_j \rangle| \leq \|\psi_j - \hat{\psi}_j\| \, \|(\hat{K} - K)\psi_j\| \\ &\leq \|\psi_j - \hat{\psi}_j\| \, \|\hat{K} - K\|_{\mathcal{S}}. \end{aligned} \tag{7.57}$$

Hence, the result in (7.56) is obtained based on the results in Theorem 2.2. Now for the second term in (7.55)

$$\begin{aligned} |\langle \psi_j, \mathrm{K}(\hat{\psi}_j - \psi_j)\rangle| &\le \|\psi_j\|\|\mathrm{K}(\hat{\psi}_j - \psi_j)\| \le \|\mathrm{K}\|\|\hat{\psi}_j - \psi_j\| \\ &= O_P(n^{-1}), \end{aligned} \tag{7.58}$$

which is also based on the results in Theorem 2.2. Hence, (7.53) is obtained by showing that, for $j = 1, \dots, d_0$,

$$|\mathrm{K}_j - (\hat{\theta}_j - \theta)| \le O_P(n^{-1}) \tag{7.59}$$

In this regard, observe that

$$\begin{aligned} |\mathrm{K}_j - (\hat{\theta}_j - \theta)| &= |\langle \psi_j, \hat{\mathrm{K}}\hat{\psi}_j\rangle - \langle \mathrm{K}\psi_j, \hat{\psi}_j\rangle - (\hat{\theta}_j - \theta_j)| \\ &= |(\hat{\theta}_j - \theta_j))(\langle \psi_j, \hat{\psi}_j\rangle - 1)| \le |\hat{\theta}_j - \theta_j||\langle \psi_j, \hat{\psi}_j\rangle - 1|, \end{aligned} \tag{7.60}$$

due to the fact that K is self-adjoint and $\mathrm{K}\psi_j = \theta_j\psi_j$, respectively. Furthermore,

$$\begin{aligned} |\langle \psi_j, \hat{\psi}_j\rangle - 1| &= \left|\int (\psi_j(u)\hat{\psi}(u) - \psi_j(u)\psi(u))\, du\right| \\ &= \left|\int \psi_j(u)(\hat{\psi}(u) - \psi_j(u))\, du\right| = |\langle \psi_j, \hat{\psi}_j - \psi_j\rangle| \le \|\hat{\psi}_j - \psi\|. \end{aligned} \tag{7.61}$$

Therefore, Theorem 2.2 leads to (7.59). This complete the proof of (7.53).

Now, we have by using (7.53)

$$\sum_{j=1}^{d_0}(\hat{\theta}_j - \theta_j) = \sum_{j=1}^{d_0}\langle \psi_j, (\hat{\mathrm{K}} - \mathrm{K})\psi_j\rangle + O_P(n^{-1}). \tag{7.62}$$

Note that $\theta_j = 0$, $\mathrm{span}\{\psi_j : j > d_0\} = \mathcal{M}^\perp$ and $\mathrm{K}\psi_j = 0$ for $j > d_0$. These and (7.62) lead to

$$\sum_{j=d_0+1}^{n} \hat{\theta}_j = \sum_{j=d_0+1}^{\infty} \langle \psi_j, (\hat{\mathrm{K}} - \mathrm{K})\psi_j\rangle + O_P(n^{-1}).$$

Moreover, by letting $\bar{\mathrm{K}} = \sum_{q=1}^{p} \tilde{\mathrm{M}}^{(q)}\mathrm{M}^{(q)*}$, we have

$$\begin{aligned} \sum_{j=d_0+1}^{n} \hat{\theta}_j &= \sum_{j=d_0+1}^{\infty} \langle \psi_j, (\tilde{K} - K)\psi_j\rangle + O_P(n^{-1}) \\ &= \sum_{j=d_0+1}^{\infty} \langle \psi_j, (\bar{K} - K)\psi_j\rangle + O_P(n^{-1}), \end{aligned} \tag{7.63}$$

where the first equality is due to the second result in (7.44) and the second equality is obtained by noting that

$$\|\tilde{\mathrm{K}} - \bar{\mathrm{K}}\|_{\mathcal{S}} \le \sum_{q=1}^{p} \|\tilde{\mathrm{M}}^{(\mathrm{q})}\tilde{\mathrm{M}}^{(\mathrm{q})^*} - \tilde{\mathrm{M}}^{(\mathrm{q})}\mathrm{M}^{(\mathrm{q})*}\|_{\mathcal{S}} = O_P(n^{-1}), \tag{7.64}$$

which is implied by Lemma 3 of Bathia et al. (2010). Since $\psi_j \in \mathcal{M}^\perp$ for $j \ge d_0 + 1$ and $\mathrm{Ker}(\bar{\mathrm{M}}^{(\mathrm{q})}) = \mathrm{Ker}(\bar{\mathrm{K}}) = \mathrm{Ker}(\mathrm{K}) = \mathcal{M}^\perp$, it holds that

$$\sum_{j=d_0+1}^{\infty} \langle \psi_j, (\bar{K} - K)\psi_j\rangle = 0. \tag{7.65}$$

Finally, by noting (7.65), the claimed result is obtained based on (7.62) and

$$
\begin{aligned}
|\langle \psi_j, (\hat{K}-K)\psi_j\rangle| &= |\langle \psi_j, (\hat{K}-K)\psi_j\rangle| \ \le \ \|\psi_j\| \ \|(\hat{K}-K)\psi_j\| \\
&\le \ \|\hat{K}-K\|_{\mathcal{S}}
\end{aligned}
$$

by which Theorem 2.2 suggests that

$$
|\langle \psi_j, (\hat{K}-K)\psi_j\rangle| \le O_P(n^{-1/2}). \tag{7.66}
$$

7.9. *Proof of Theorem 3.2(i)*

Let us observe firstly that

$$
\begin{aligned}
IC(d) - IC(d_0) &= \left\{\hat{S}^{(d)} - \hat{S}^{(d_0)}\right\} - (d-d_0)P_n \\
&= \left\{\hat{S}^{(d)} - S^{(d)}\right\} - \left\{\hat{S}^{(d_0)} - S^{(d_0)}\right\} + \left\{S^{(d)} - S^{(d_0)}\right\} - (d-d_0)P_n.
\end{aligned}
$$

When $d > d_0$,

$$
\begin{aligned}
\left\{\hat{S}^{(d)} - S^{(d)}\right\} - \left\{\hat{S}^{(d_0)} - S^{(d_0)}\right\} &= \sum_{j=1}^{d_0}(\hat{\theta}_j - \theta_j) + \sum_{j=(d_0+1)}^{d}(\hat{\theta}_j - \theta_j) - \sum_{j=1}^{d_0}(\hat{\theta}_j - \theta_j) \\
&= (d-d_0)O_P(n^{-1/2}) \qquad (7.67)
\end{aligned}
$$

by using Theorem 3.3, and

$$
\begin{aligned}
IC(d) - IC(d_0) &= \left\{S^{(d)} - S^{(d_0)}\right\} + (d-d_0)O_P(n^{-1/2}) - (d-d_0)P_n \\
&= (d-d_0)O_P(n^{-1/2}) - (d-d_0)P_n < 0, \qquad (7.68)
\end{aligned}
$$

where the above inequality holds by the condition (b) of the the theorem. Furthermore, when $d < d_0$,

$$
\begin{aligned}
\left\{\hat{S}^{(d)} - S^{(d)}\right\} - \left\{\hat{S}^{(d_0)} - S^{(d_0)}\right\} &= \sum_{j=1}^{d}(\hat{\theta}_j - \theta_j) - \sum_{j=1}^{d}(\hat{\theta}_j - \theta_j) - \sum_{j=(d_0+1)}^{d}(\hat{\theta}_j - \theta_j) \\
&= (d-d_0)O_P(n^{-1/2}) \qquad (7.69)
\end{aligned}
$$

also by using Theorem 3.3, and

$$
IC(d) - IC(d_0) \ = \ \left\{S^{(d)} - S^{(d_0)}\right\} + (d-d_0)O_P(n^{-1/2}) - (d-d_0)P_n < 0, \tag{7.70}
$$

where the inequality holds almost surely for sufficiently large n. Only when $d = d_0$ that $IC(d) - IC(d_0) = 0$. Accordingly, $\hat{d}$ that maximizes $IC(d)$ converges in probability to d_0 as $n \to \infty$.

7.10. Proof of Theorem 3.2(ii)

Let us observe firstly that $\left\{S^{(d)} - S^{(d_0)}\right\} = -\sum_{j=d+1}^{d_0} \theta_j$ for $d < d_0$, $\left\{S^{(d)} - S^{(d_0)}\right\} = 0$ for $d = d_0$, and $\left\{S^{(d)} - S^{(d_0)}\right\} = 0$ for $d > d_0$.

Now, let us introduce $d_0' > d_0$. Then $\left\{S^{(d)} - S^{(d_0')}\right\} = -\left(\sum_{j=d+1}^{d_0} \theta_j + \sum_{j=d+1}^{d_0'} \theta_j\right)$ for $d < d_0$, $\left\{S^{(d)} - S^{(d_0')}\right\} = -\sum_{j=d_0+1}^{d_0'} \theta_j$ for $d = d_0$, and $\left\{S^{(d)} - S^{(d_0')}\right\} = -\sum_{j=d_0+1}^{d_0'} \theta_j$ for $d > d_0$.

Let us also introduce $d' > d$. Then $\left\{S^{(d')} - S^{(d_0')}\right\} = -\sum_{j=d'+1}^{d_0'} \theta_j$ for $d' < d_0'$, $\left\{S^{(d')} - S^{(d_0')}\right\} = 0$ for $d' = d_0'$, and $\left\{S^{(d')} - S^{(d_0')}\right\} = 0$ for $d' > d_0'$.

The above two points suggest therefore that $S^{(d')} > S^{(d)}$. Furthermore, we have by Theorem 3.3

$$\begin{aligned} IC(d) &= \hat{S}^{(d)} + dP_n = (\hat{S}^{(d)} - S^{(d)}) + S^{(d)} + dP_n \\ &= S^{(d)} + dP_n + O_P(n^{-1/2}) \end{aligned} \tag{7.71}$$

and

$$\begin{aligned} IC(d') &= \hat{S}^{(d')} + d'P_n = (\hat{S}^{(d')} - S^{(d')}) + S^{(d')} + d'P_n \\ &= S^{(d')} + d'P_n + O_P(n^{-1/2}), \end{aligned} \tag{7.72}$$

which suggest that $IC(d') > IC(d)$. Hence, when d_0 increases to d_0', i.e. $d_0' > d_0$, $d' > d$ is selected. In this regard, Theorem 3.3 suggests therefore that

$$\lim_{n\to\infty} \mathrm{Prob}(\hat{d}' = d_0') = 1. \tag{7.73}$$

This holds for the case in which $d_0 = d_n$ is considered to be a function of n and d_n tend to infinity.

Nonetheless, d_n must not converge to infinity faster than $n^{1/2}$. To see this, observe that (7.70) in the proof of Theorem 3.2(i) can be re-written as

$$IC(d) - IC(d_0) = \left\{S^{(d)} - S^{(d_0)}\right\} + (d - d_0)O_P(n^{-1/2}) + (d_0 - d)P_n < 0. \tag{7.74}$$

Therefore, we are able to ensure that such an inequality hold for the case in which $d_0 = d_n$ tends to infinity faster $n^{1/2}$.

8. References

AIELLI, G. P. (2013). Dynamic conditional correlation: On properties and estimation. *J. Bus. Econom. Statist.* 31 282–299. MR3173682.

ANG, A. and CHEN, J. (2002). Asymmetric correlations of equity portfolios. *J. Financ. Econ.* 63 443–494.

ASLANIDIS, N. and CASAS, I. (2013). Noparametric correlation models for portfolio allocation. *J. Bank. Financ.* 37 2268–2283.

BAI, J. and NG, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* 70 191–221. MR1926259

BATHIA, N., YAO, Q. and ZIEGELMANN, F. (2010). Identifying the finite dimensionality of curve time series. *Ann. Statist.* 38 3352-3386.

BENKO, M., HARDLE, W. and KNEIP, A. (2009). Common functional principal components. *Ann. Statist.* 37 1–34. MR2488343

CAI, T. and HALL, P. (2006). Prediction in Functional Linear Regression. *Ann. Statisti.* 34, 2159–2179.

CAPPIELLO, L., ENGLE, R. F. and SHEPPARD, K. (2006) Asymmetric dynamics in the correlations of global equity and bond returns. *J. Fin. Econometrics* 4 537-572.

DALLA, V., GIRAITIS, L. and PHILLIPS, PCB. (2015). Testing mean stability of heteroskedastic time series, *Working Paper.*

ENGEL, C., MARK, N. C. and WEST, K. (2015) Factor Model Forecasts of Exchange Rates, *Econometric Rev.* 34 32-55.

ENGLE, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *J. Bus. Econom. Statist.* 20 339–350. MR1939905

FAN, J., and YAO, Q. (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods* Springer, New York.

GREENAWAY-MCGREVY, R., MARK, N. C. and DONGGYU, S. and WU, J. L. (2012) Exchange rates as exchange rate common factors. *Hong Kong Institute for Monetary Research* 212012.

HAFNER, C. M., VAN DIJK D. and FRANSES, P. H. (2006) Semi-Parametric modeling of correlation dynamics. In *Econometric Analysis of Financial and Economic Time Series* (D. Terrell, T. B. Fomby, eds.) 59-103.

HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B.* 68, 109–126.

HALL, P. and VIAL, C. (2006). Assessing the finite dimensionality of functional data. *J. R. Stat. Soc. Ser. B.* 68 689–705.

HANSEN, B. E. (2003) Uniform convergence rates for kernel estimation with dependent data. *Econometric Thoe.* 24 726–748.

HAYS, S., SHEN, H. and HUANG, J.Z. (2012). Functional dynamic factor models with application to yield curve forecasting. *Ann. App. Statist.* 6 870-894.

HERSKOVIC, B., KELLY, B., LUSTIG, H. and NIEUWERBURGH, S. V. (2016) The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *J. Fin. Econ.* 119 249-283.

HYNDMAN, R. J. and ULLAH, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Comput. Statist. Data Anal.* 51 4942–4956. MR2364551

JIANG, H., SAART, P. W. and XIA Y. (2016). Asymmetric conditional correlations in stock returns. *Ann. App. Statist.* 10 989-1018.

KASCH, M. and CAPORIN, M. (2013) Volatility threshold dynamic conditional correlations: An international analysis. *J. Fin. Econometrics* 11 706-742.

KÖBER, L., LINTON, O., and VOGT, M. (2015) A semiparametric model for heterogeneous panel data with fixed effects. *J. Econometrics* 188 327-345.

LI, Y., WANG, N. and CARROLL, R. J. (2013). Selecting the number of principal components in functional data. *J. Amer. Statist. Assoc.* 108 1284–1294. MR3174708

LIEBL, D. (2013) Modeling and forecasting electricity spot prices: A functional data perspective. *Ann. App. Statist.* 7 1562-1592.

LEE, A. J. (1990). U-Statistics. Dekker, New York.

LÜTKEPOHL, H. (2005). *New introduction to multivariate time series analysis.* Springer, New York.

MERCER, J. (1909) Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society London (A).* 209 415-446

MÜLLER, H., RITUPARNA, S. and ULRICH, S. (2011). Functional data analysis for volatility. *J. Econometrics.* 165 233-245.

PORTER, D. and STIRLING, D.G. (1990). *Integral Equations: a Practical Treatment from Spectral Theory to Applications.* Cambridge Texts in Applied Mathematics, Cambridge

PELLETIER, D. (2006). Regime switching for dynamic correlations. *J. Econometrics* 131 445–473. MR2276007

RENAULT, E., HEIJDEN, T. V. D. and WERKER, J. M. (2017). APT with idiosyncratic variance factors. *Working Paper.*

SILVENNOINEN, A. and TERÄSVIRTA, T. (2015). Modeling conditional correlations of asset returns: A smooth transition approach. *Econometric Rev.* 34 174–197. MR3268917

VERDELHAN, A. (2018) The share of systematic variation in bilateral exchange rates. *J. Fin.* 73 375-418.

VAN DIJK, D., MUNANDAR, H. and HAFNER, C. (2006) The Euro introduction and non-Euro currencies. *medium econometrische toepassingen.* 14 30-36.

WANG, L. (2008) Karhunen-Loeve Expansions and their applications. *Unpublished doctoral dissertation.* London School of Economics and Political Science, London.

YAO, F., MULLER, H.G., and WANG, J.L. (2005), Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* 100, 577–590.

ZIVOT, E. and WANG, J. (2013) *Modeling Financial Time Series with S-PLUS* Springer, New York.

Fig 1: Percentages of accurate selection in Table 1 plotted by m

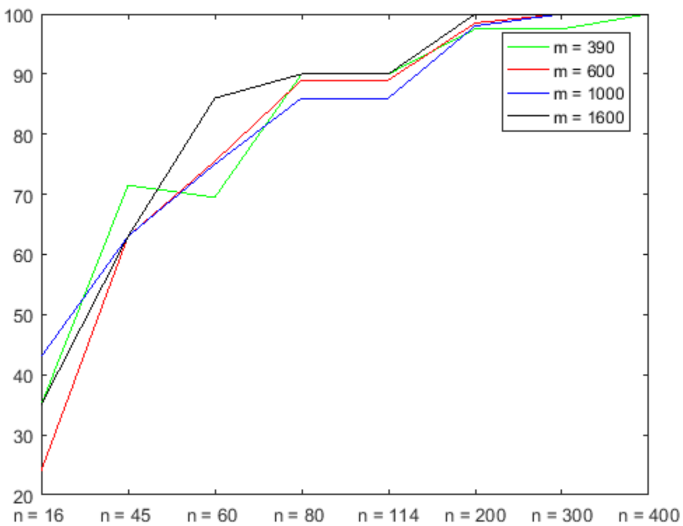


Fig 2: Medians of the D measure in Table 2 plotted by m

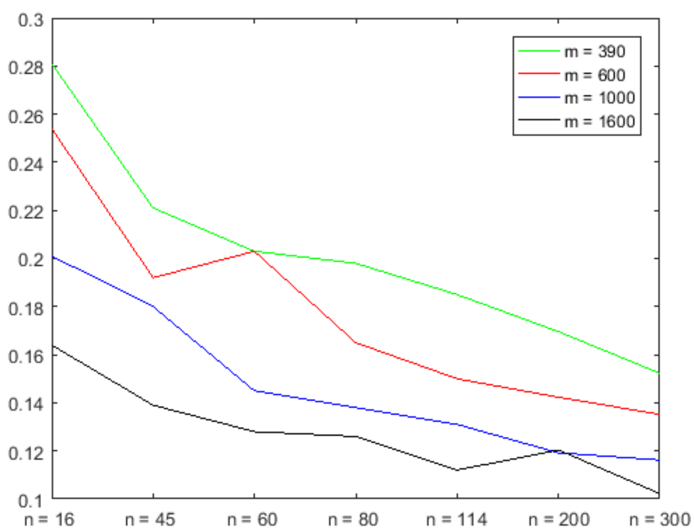
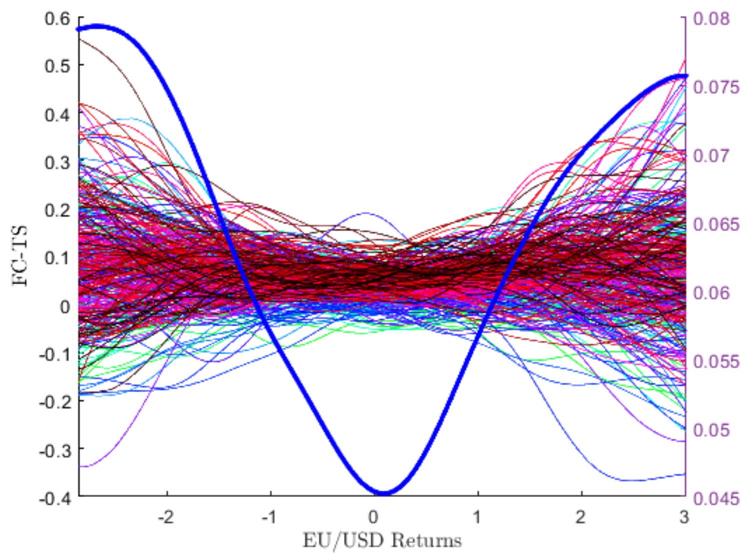
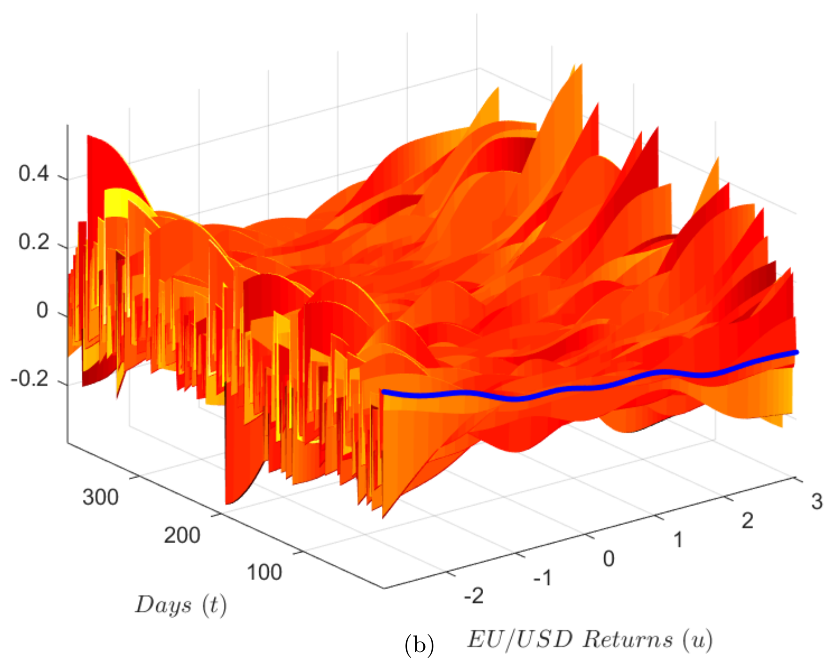


Fig 3: 2D and 3D plots of functional correlations time series (FC-TS) of the British Pound and Swiss Franc, i.e. $\hat{\rho}_{chf,1}(u), \ldots, \hat{\rho}_{chf,n}(u)$



(a)



(b)

Fig 4: Information criterion $1 \leq d \leq 10$ for selecting number of eigenfunctions based on FC-TS for the British Pound and Swiss Franc

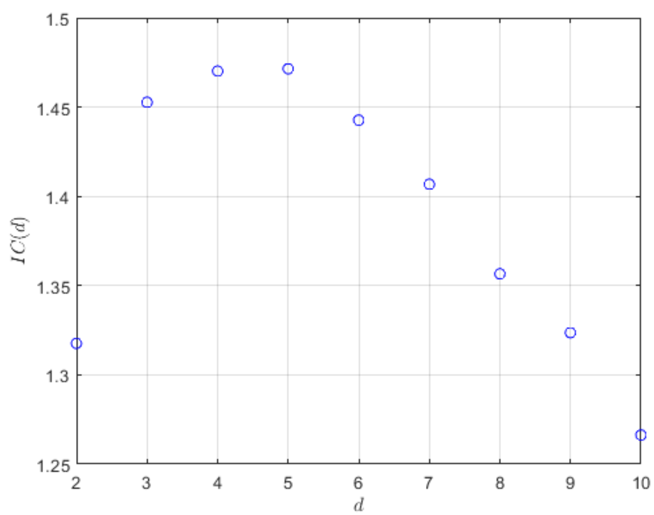


Fig 5: Autocorrelation functions for the estimated loading time series, $\hat{\eta}_{chf,t,1}, \ldots, \hat{\eta}_{chf,t,6}$, based on FC-TS for the British Pound and Swiss Franc

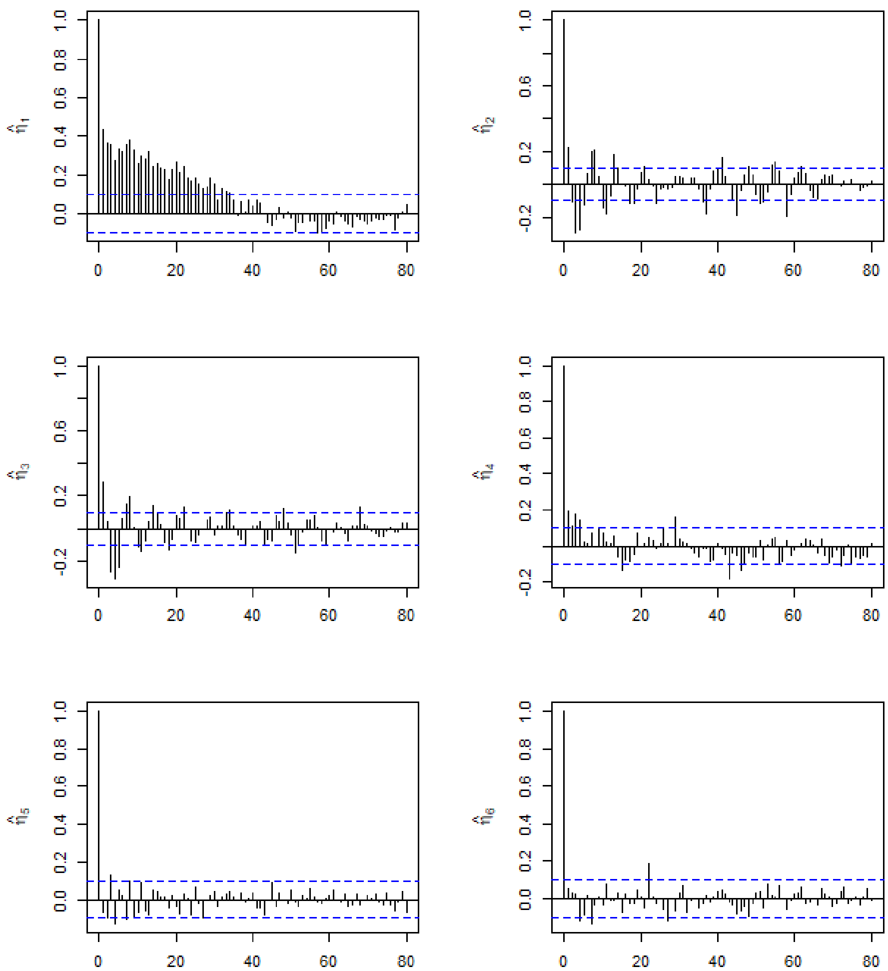


Fig 6: Estimated eigenfunctions corresponding to first five eigenvalues based on FC-TS of the British Pound and Swiss Franc

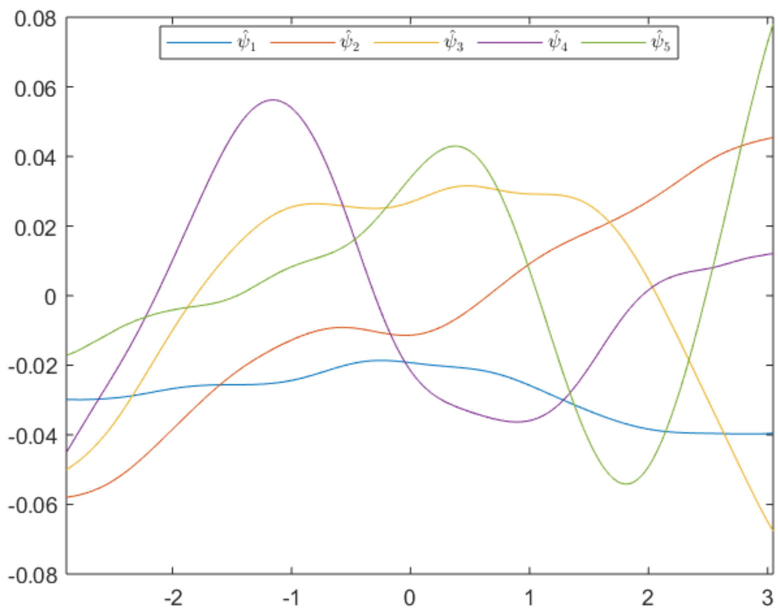


Fig 7: Fitting or in-sample forecasts $(\hat{\rho}^{(5)}_{chf,t}(u))$ [black], estimated FC-TS of the British Pound and Swiss Franc $(\hat{\rho}_{chf,t}(u))$ [red] and estimated mean correlation function $(\hat{\varrho}_{chf}(u))$ [blue]

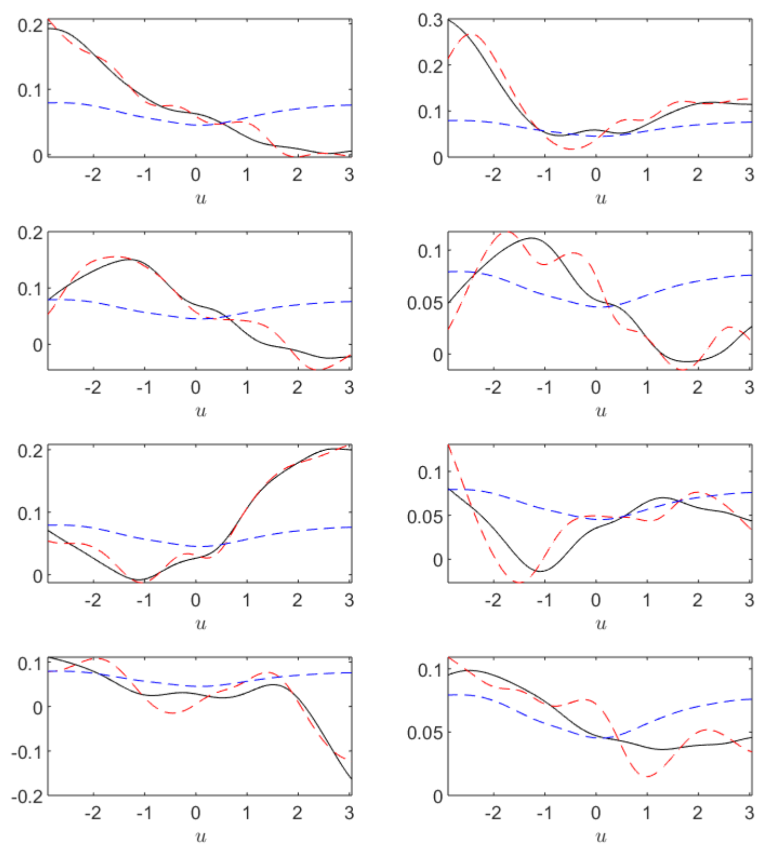


Fig 8: Percentage of the auto-covariance being explained based on FC-TS of the British Pound and Swiss Franc

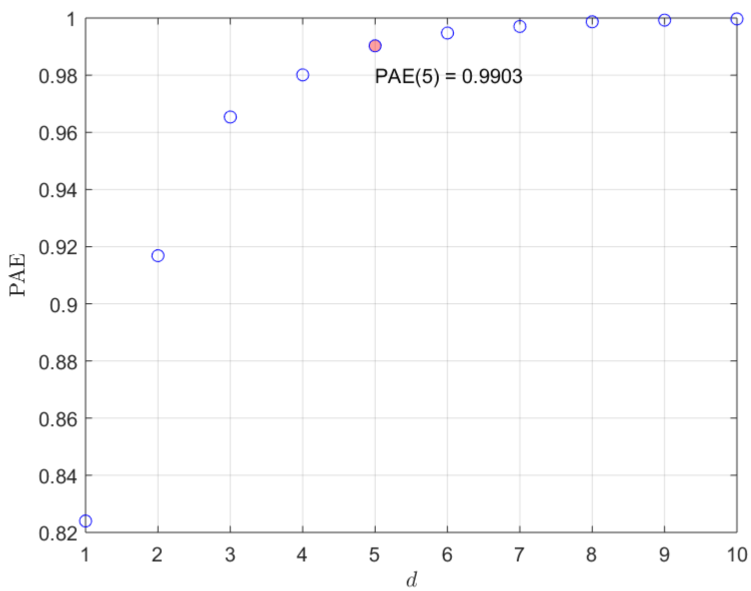
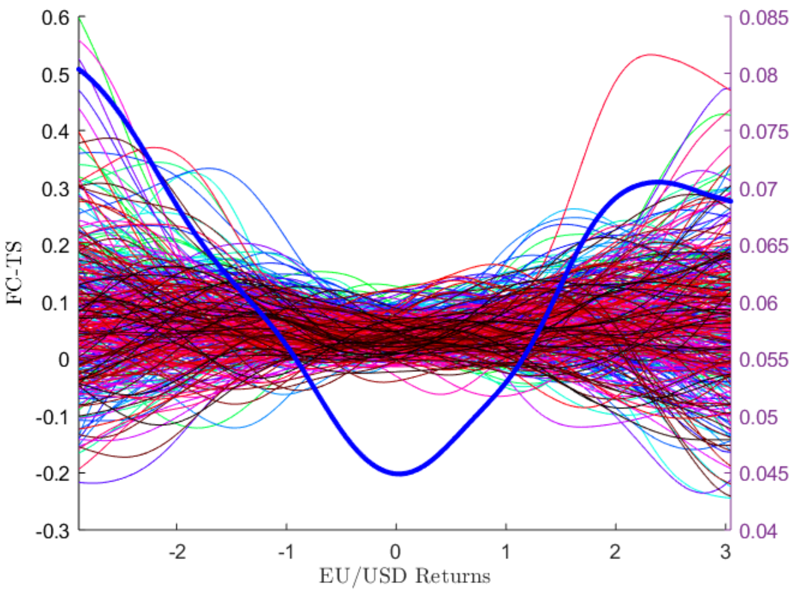
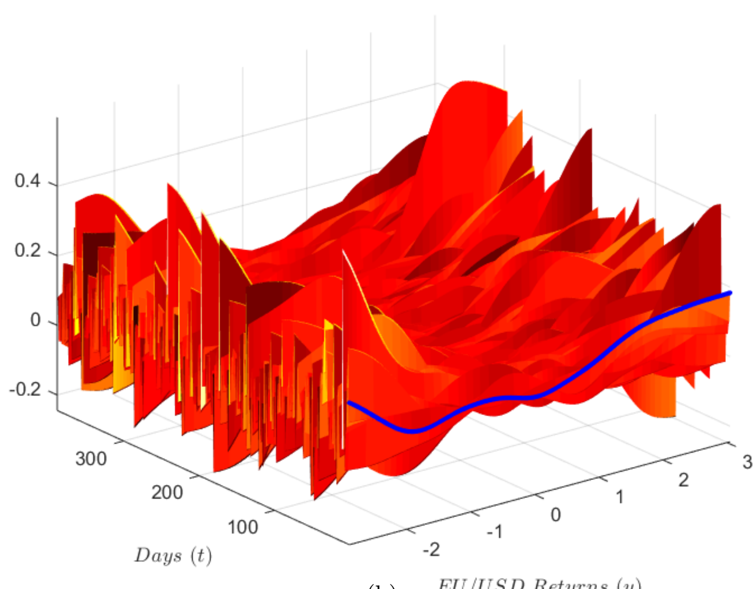


Fig 9: 2D and 3D plots of FC-TS for the British Pound and Swedish Krona, i.e. $\hat{\rho}_{sek,1}(u), \ldots, \hat{\rho}_{sek,n}(u)$



(a)



(b)

Fig 10: Information criterion $1 \leq d \leq 10$ for selecting number of eigenfunctions based on FC-TS for the British Pound and Swedish Krona

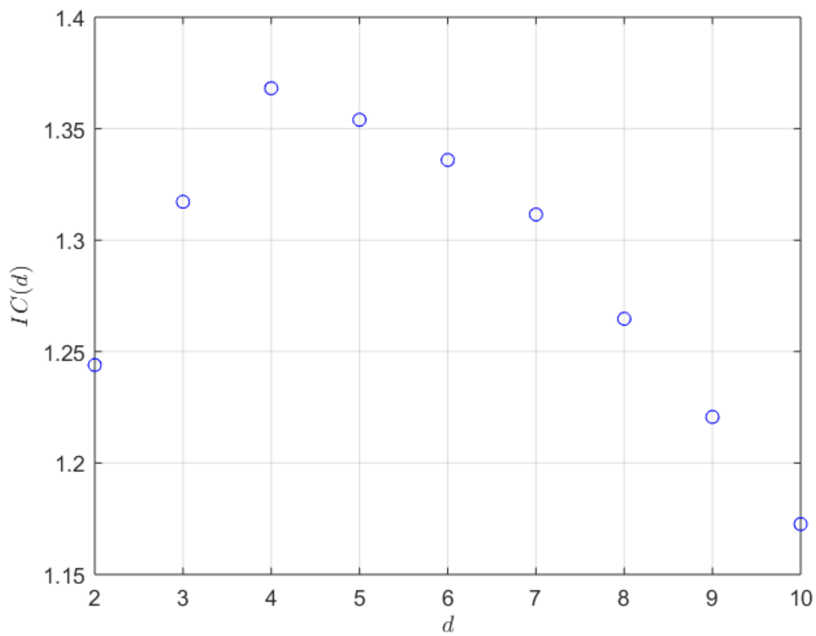


Fig 11: Autocorrelation functions for the estimated loading time series, $\hat{\eta}_{sek,t,1}, \ldots, \hat{\eta}_{sek,t,6}$ based on FC-TS of the British Pound and Swedish Krona

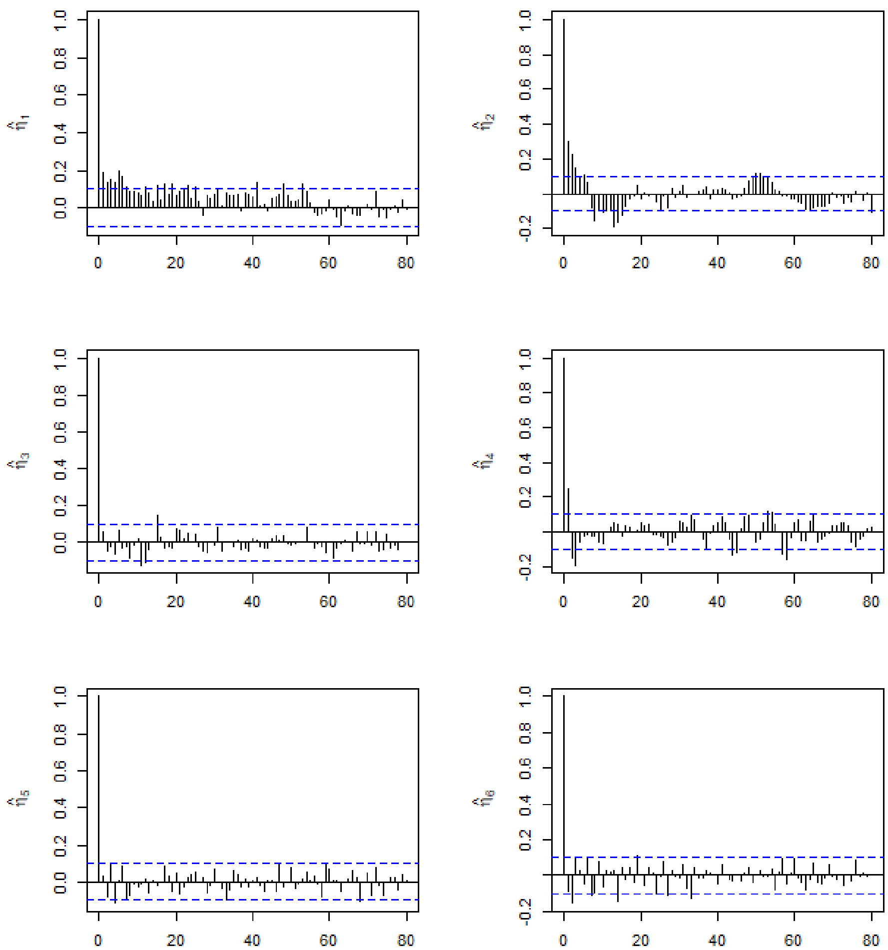


Fig 12: Information criterion $1 \leq d \leq 10$ for selecting number of eigenfunctions based on FC-TS of the British Pound and Swedish Krona

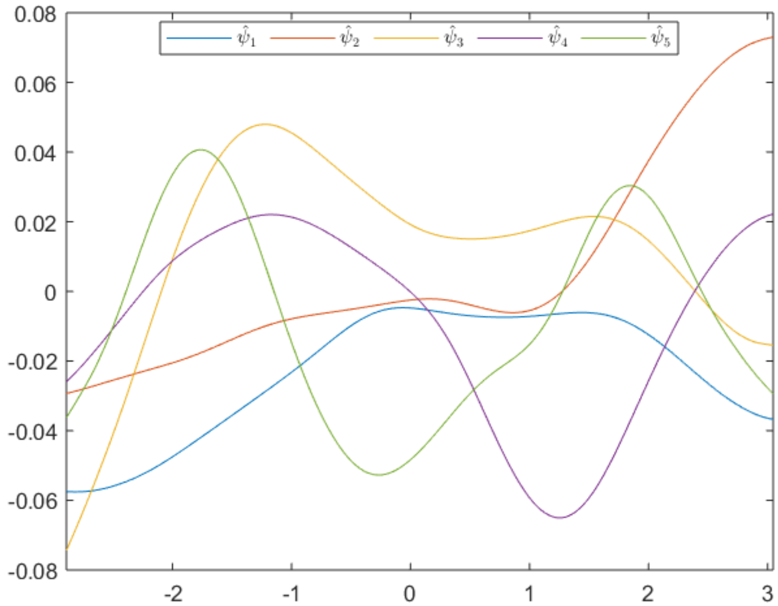


Fig 13: Fitting or in-sample forecasts $(\hat{\rho}^{(5)}_{sek,t}(u))$ [black], estimated FC-TS of the British Pound and Swedish Krona $(\hat{\rho}_{sek,t}(u))$ [red] and estimated mean correlation function $(\hat{\varrho}_{sek}(u))$ [blue]

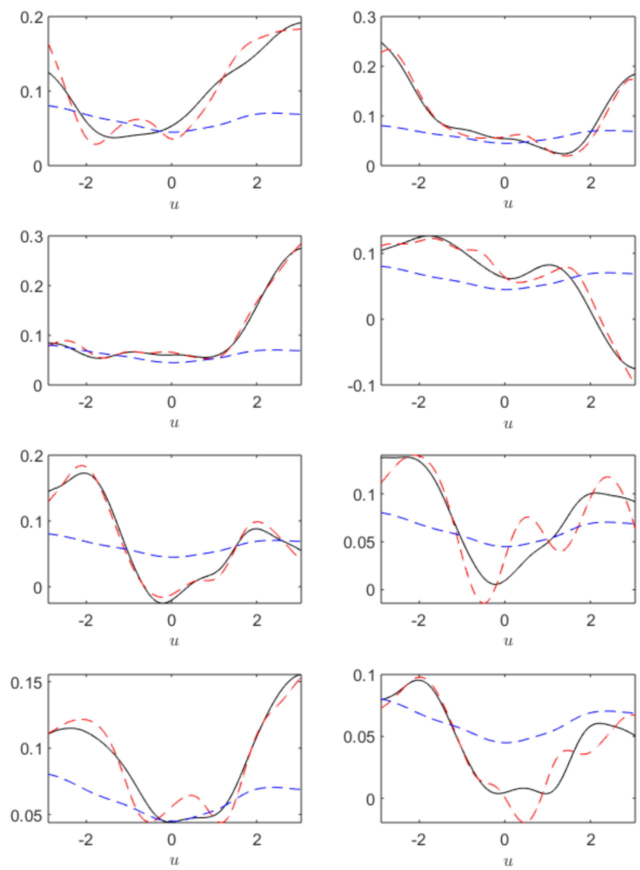


Fig 14: Percentage of the auto-covariance being explained based on FC-TS of the British Pound and Swedish Krona

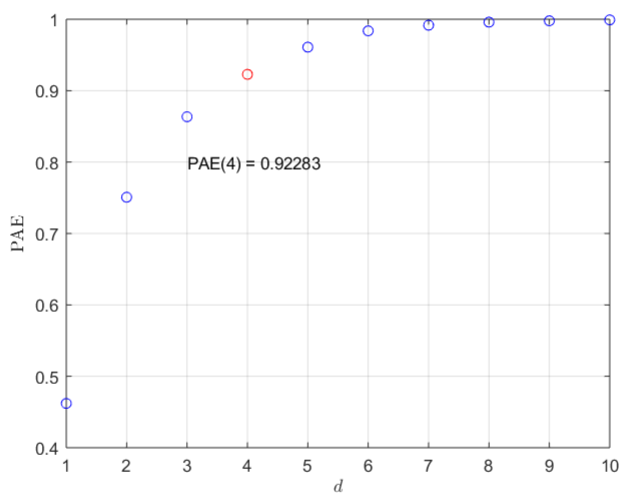
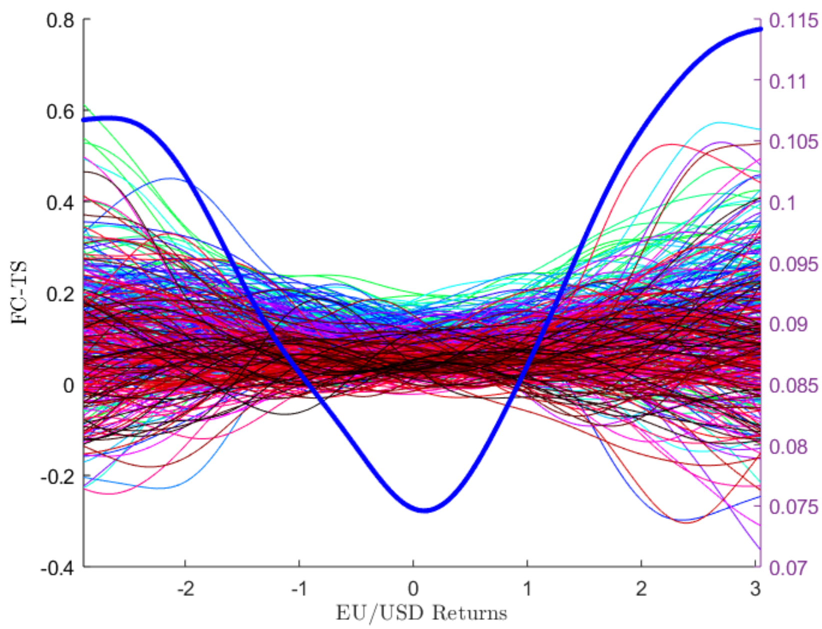
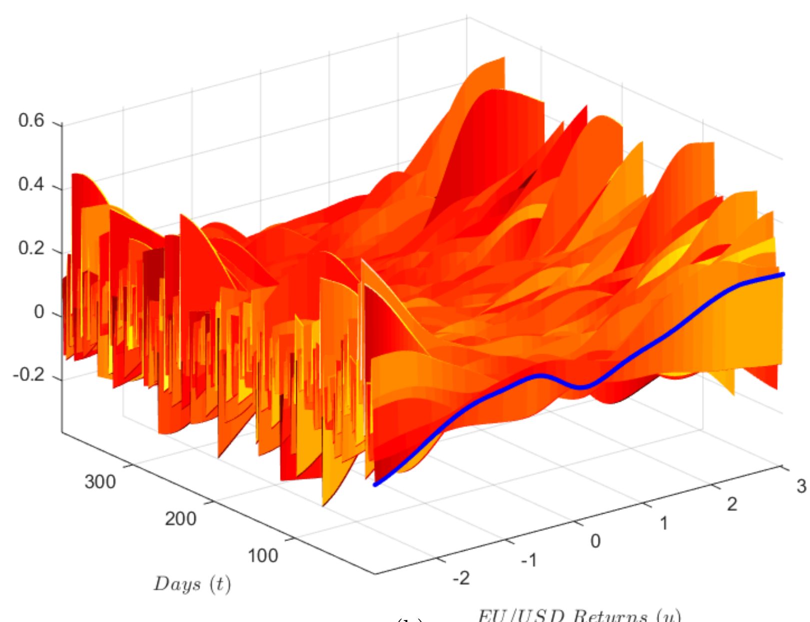


Fig 15: 2D and 3D plots of FC-TS of the British Pound and Norwegian Krone, i.e. $\hat{\rho}_{nok,1}(u), \ldots, \hat{\rho}_{nok,n}(u)$



(a)



(b)

Fig 16: Information criterion $1 \leq d \leq 10$ for selecting number of eigenfunctions based on FC-TS of the British Pound and Norwegian Krone

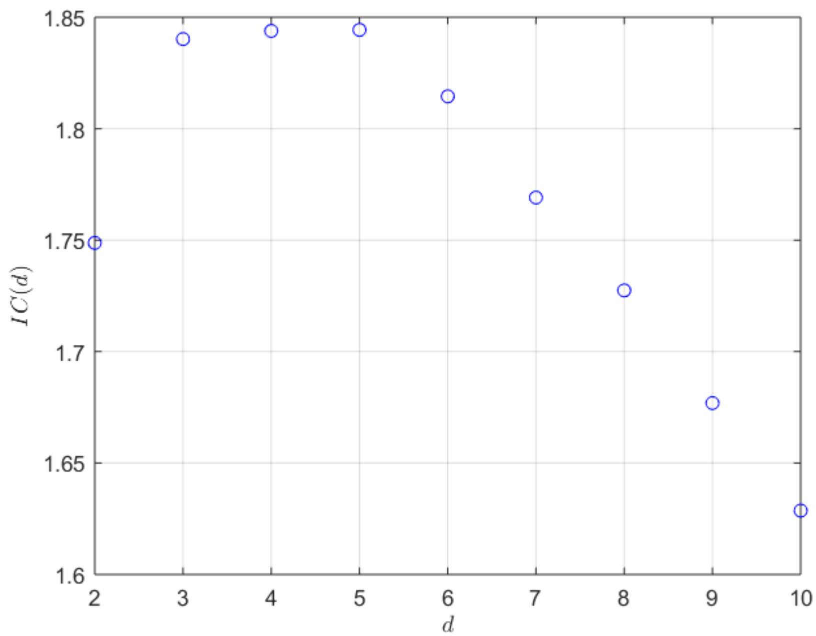


Fig 17: Autocorrelation functions for the estimated loading time series, $\hat{\eta}_{nok,t,1}, \ldots, \hat{\eta}_{nok,t,6}$ based on FC-TS of the British Pound and Norwegian Krone

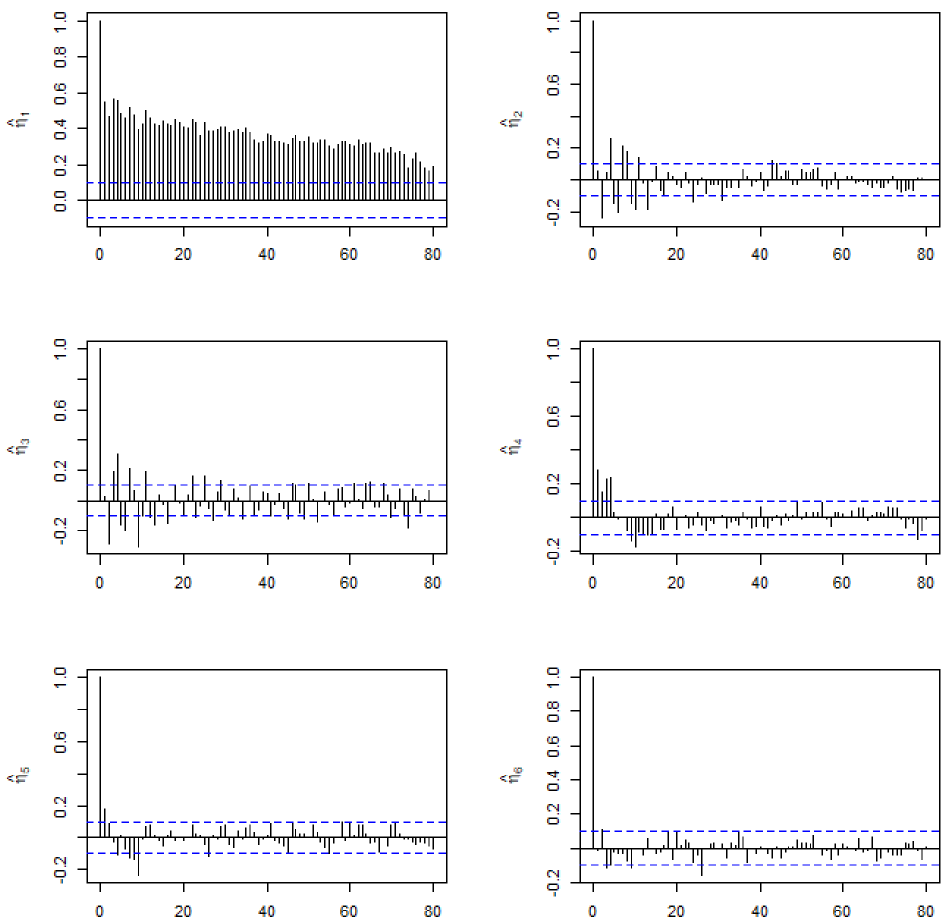


Fig 18: Information criterion $1 \leq d \leq 10$ for selecting number of eigenfunctions based on FC-TS of the British Pound and Norwegian Krone

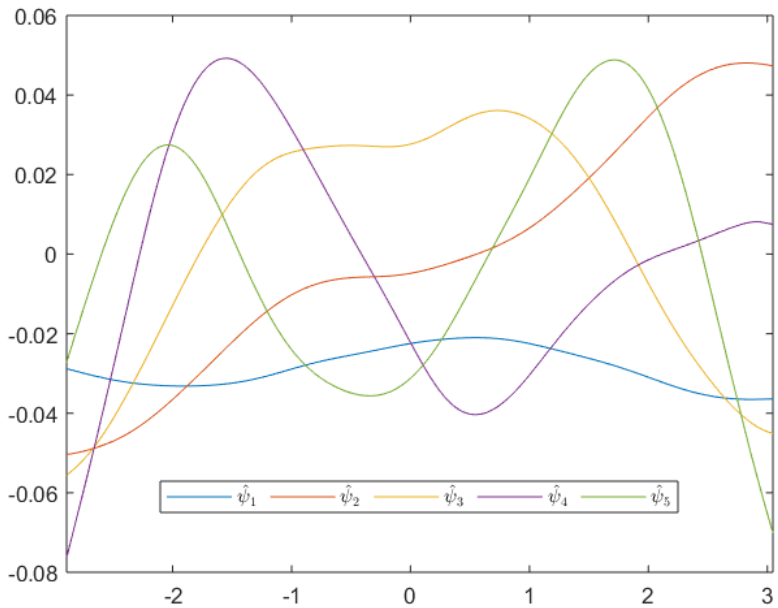


Fig 19: Fitting or in-sample forecasts $(\hat{\rho}^{(5)}_{nok,t}(u))$ [black], estimated FC-TS of the British Pound and Norwegian Krone $(\hat{\rho}_{nok,t}(u))$ [red] and estimated mean correlation function $(\hat{\varrho}_{nok}(u))$ [blue]

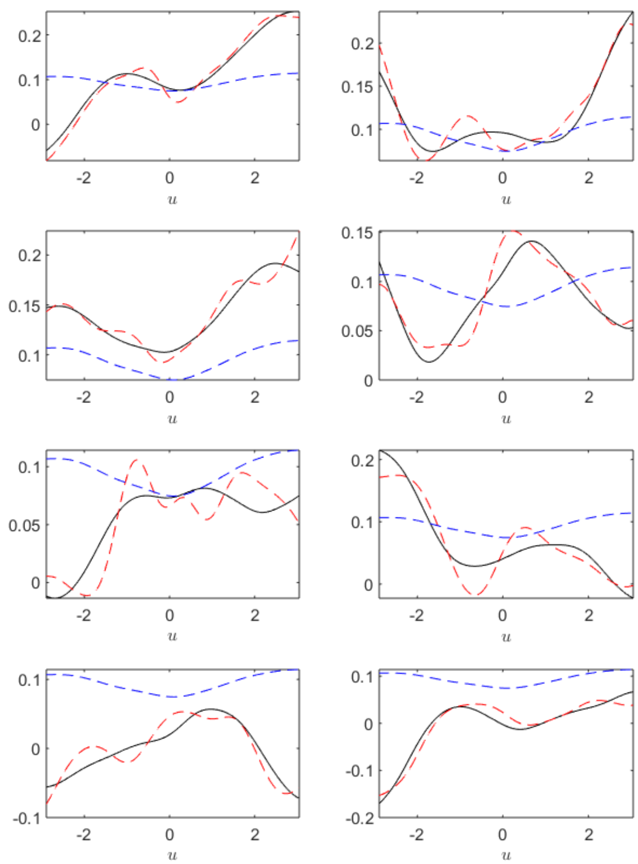


Fig 20: Percentage of the autocovariance of FC-TS of the British Pound and Norwegian Krone being explained