Inférence asymptotique pour des processus stationnaires fonctionnels

Clément Cerovecki

Thèse présentée en vue de l'obtention du titre de docteur en sciences de l'Université Libre de Bruxelles et de Université de Lille. Soutenue en privé le 22 mai 2018 et en public le 7 juin 2018.

Composition du jury :

Christian Francq	Université de Lille	Directeur
Siegfried Hörmann	Université Libre de Bruxelles	Directeur
Christine De Mol	Université Libre de Bruxelles	Présidente
Jean-Michel Zakoïan	Université de Lille	Examinateur
Irene Gijbels	Katholieke Universiteit Leuven	Rapporteuse
Piotr Kokoszka	Colorado State University	Rapporteur

Année académique 2017–2018

Remerciements

Je tiens tout particulièrement à remercier mes deux superviseurs. Siegfried Hörmann à l'ULB, avec qui j'ai vraiment fais mes premiers pas dans la recherche, dès le master et pendant les premiers mois du doctorat, puis qui m'a toujours guidé avec beaucoup d'enthousiasme et un peu de folie, notamment lors des longues séances d'exploration au tableau. Christian Francq à Lille 3, de m'avoir fait confiance en acceptant de me superviser sans trop me connaître. Puis également Jean–Michel Zakoïan qui, avec Christian, m'a si bien accueilli lors de mes séjours à Paris. Tous deux prenaient le temps de réflechir avec moi sur chaque détail et c'était toujours un grand plaisir de travailler ensemble.

Je remercie le Fond de la Recherche Scientifique, FNRS, qui a financé ces 4 années de doctorat via une bourse FRIA.

Je remercie également Irène Gijbels, Christine de Mol et Piotr Kokoszka d'avoir accepté de faire partie de mon jury de thèse.

Enfin, je remercie Vaidotas Characiejus avec qui j'ai eu le plaisir de collaborer ces derniers mois. Les professeurs du département, avec notamment Davy pour l'excellente introduction au statistiques en Master. Puis Malou, et son précieux soutien pour dompter la paperasse administrative. Les doctorants du 9ème étage, Rémy, Sarah, Gilles, Mohammed, Bastien, Pierre, Julien. Monsieur Montada, le génial prof de secondaire qui m'a fait découvrir les maths. Et bien sûr, Laurence, sans qui je n'aurais pas eu le courage de faire tout ça.

Contents

Résumé de la thèse

Introduction

Cette thèse aborde divers problèmes concernant l'*analyse des séries temporelles fonctionnelles*. Comme son nom l'indique, cette discipline se trouve à l'intersection de deux autres, à savoir l'analyse des *données fonctionnelles*, et celle des *séries temporelles*. En pratique, de nombreuses données peuvent s'interpréter comme des *fonctions* ou des *courbes*, par exemple la Figure 1 représente un échantillon issu d'une étude kinésiologique. Bien que le développement des statistiques fonctionnelles ait été entamé par Karhunen (1947), il faut attendre le développement informatique de ces dernières décennies pour voir apparaître de véritables méthodes de traitement des données fonctionnelles. On considère alors un échantillon $X_1, \dots, X_n$, où l'on suppose que chaque observation X_t, est la réalisation d'une certaine fonction aléatoire $u \mapsto X_t(u)$. Théoriquement, la principale difficulté est qu'on dispose d'un nombre fini d'observations, à valeurs dans un espace de dimension infinie. En pratique, on observe seulement une discrétisation de chaque courbe $X_t(u_1), \dots, X_t(u_N)$, donc un vecteur, qu'il serait tentant de traiter comme en analyse multivariée. Or cette approche s'avère inadéquate, car il ne s'agit pas de vecteurs aléatoires quelconques. Notamment la continuité des courbes induit une forte corrélation entre deux évaluations voisines $X_t(u_i)$ et $X_t(u_{i+1})$. Aussi, les points d'échantillonnage u_i ne sont pas nécessairement synchronisés pour toutes les observations. Une véritable approche fonctionnelle est donc préférable. Concrètement, on représente les observations dans une certaine base $X_t(u) = \sum_{k=1}^{p} \mathbf{x}_{t,k}\varphi_k(u)$, pour tout $u \in [0,1]$. Cette étape s'appelle le lissage, voir Figure 1, on travaille alors avec les coefficients $\mathbf{x}_1, \dots, \mathbf{x}_n$ dans $\mathbb{R}^p$, mais où p n'est pas fixé préalablement, comme en analyse multivariée. Pour plus de détails sur le lissage des données fonctionnelles, on peut consulter Ramsay et Silverman (2006).

Dans notre cas, on considère des phénomènes continus observés au cour du *temps*, qui ont été segmentés. Par exemple si l'on observe des données de pollutions, il est naturel de segmenter au niveau des jours, car il y a un certain nombre de phénomènes quotidients sous-jacents, comme le trafic des voitures, idem pour un indice financier, voir Figures 2 et 3. On obtient alors une *série temporelle fonctionnelle* $(X_t)_{t\in\mathbb{Z}}$, c'est–à–dire un processus stochastique discret à valeurs dans un *espace fonctionnel* H. Le phénomène initial étant continu, on ne peut pas supposer que les observations soient *indépendantes* entre elles, d'où la terminologie *série temporelle*. Par la suite on supposera que H est un espace de Hilbert muni du produit scalaire $\langle \cdot, \cdot \rangle$ et de la norme correspondante $\|\cdot\|$. Les notions usuelles de probabilités sont valables dans les espaces de Hilbert, voir par exemple dans Bosq (2000). On supposera également que $(X_t)_{t\in\mathbb{Z}}$ est stationnaire, c'est–à–dire que $(X_{t_1}, \dots, X_{t_k}) \overset{d}{=} (X_{t_1+h}, \dots, X_{t_k+h})$, pour tout $t_1, \dots, t_k, h \in \mathbb{Z}$ et $k \geq 1$. Si $E\|X_t\| < \infty$ on supposera aussi que le processus est centré c'est–à–dire que $EX = 0$. Pour une série stationnaire, on définit les *opérateurs de covariances* par $C_h = E[X_h \otimes X_0]$, où pour tout $x, y \in H$, on a $x \otimes y = x\langle \cdot, y \rangle$. Un outil théorique important est la représentation de *Karhunen–Loève*,

$$X_t = \sum_{k=1}^{\infty} \langle X_t, v_k \rangle v_k,$$

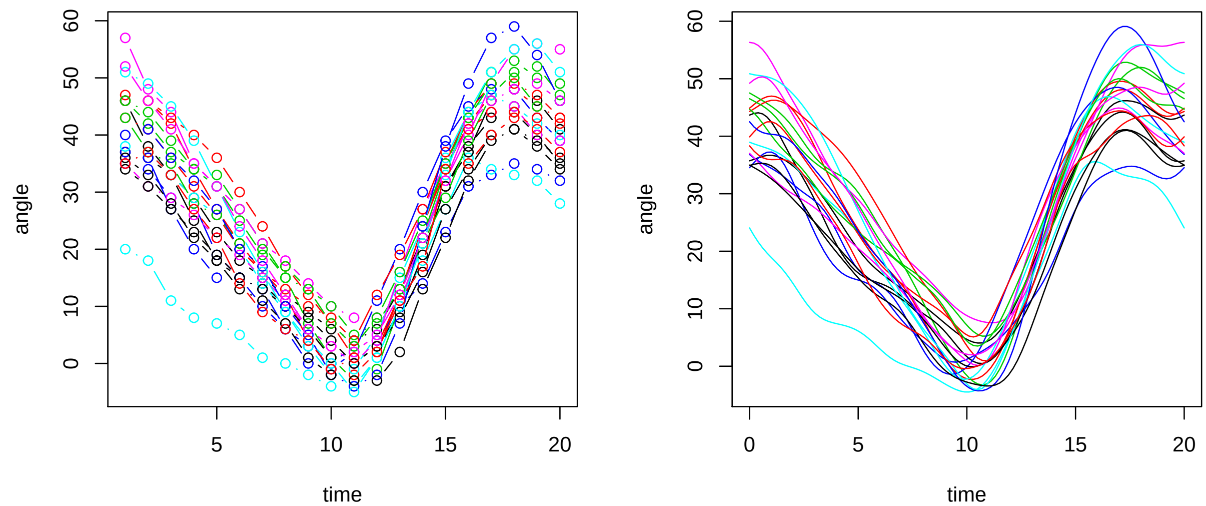

Figure 1: Angle de la hanche pendant un cycle de pas chez 20 enfants, mesurés à 21 temps réguliers. Données discrète initiales (gauche) et données lissées (droite).

où $(v_k)_{k\geq 1}$ est la base orthonormale de H, formée par les fonctions propres de l'opérateur C_0, associées aux valeurs propres $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. Un ouvrage de référence sur l'inférence pour les séries temporelles fonctionnelles est Horváth et Kokoszka (2012)

Normalité asymptotique de la transformée de Fourier

Dans un premier temps, nous nous sommes intéressés à l'analyse spectrale des séries temporelles fonctionnelles. Il s'agit d'aborder ces processus temporels via le domaine des fréquences. Pour ce faire, on définit la *transformée de Fourier discrète*

$$\mathcal{X}_n(\omega) = n^{-1/2} \sum_{t=1}^{n} X_t e^{-\mathrm{i}t\omega}, \quad \omega \in [0, \pi[, \tag{0.1}$$

associée à un échantillon $X_1, \ldots, X_n$, ainsi que l'*opérateur de densité spectrale*

$$\mathcal{F}_\omega := \sum_{h\in\mathbb{Z}} C_h e^{-\mathrm{i}h\omega}, \quad \omega \in [0, \pi[, \tag{0.2}$$

associé à une série $(X_t)_{t\in\mathbb{Z}}$. Ce dernier contient la même information que l'ensemble des opérateurs de covariances $(C_h)_{h\in\mathbb{Z}}$. Même si les observations sont réelles, on remarque qu'ici, il est nécessaire de travailler dans un espace de Hilbert complexe, on suppose donc désormais que $H = H_0 \oplus \mathrm{i}H_0$, où H_0 est un espace de Hilbert réel. La principale différence étant la caractérisation de la variance d'un élément aléatoire Z dans H. En effet, celle-ci nécessite deux opérateurs: l'opérateur de *variance* $\Gamma = E[Z \otimes Z]$ et l'opérateur de *relation* $R = E[Z \otimes \overline{Z}]$. Par exemple, une variable Gaussienne dans H est caractérisée par sa moyenne $\mu = E[Z]$ et ces deux opérateurs, et on note $Z \sim \mathcal{CN}_H(\mu, \Gamma, R)$. Plus précisément, on étudie le comportement

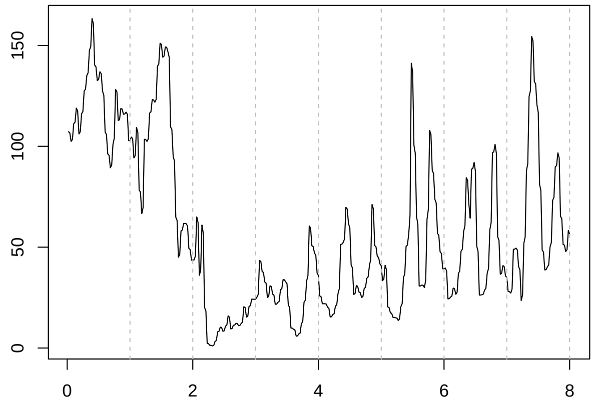

Figure 2: PM_{10} ($\mu g/m^3$) Graz

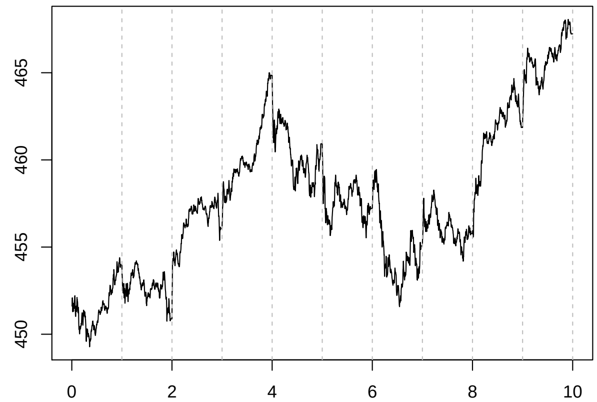

Figure 3: Indice de marché S&P

asymptotique de la transformée de Fourier discrète et l'objectif est de travailler dans un cadre aussi général que possible. Par exemple, supposons que $X_1, \ldots, X_n$ est un échantillon Gaussien i.i.d. à valeurs réelles. On a sait que $\mathcal{X}_n(\omega) \sim \mathcal{CN}_H(0, \mathcal{F}_\omega, 0)$. D'autre part si on considère deux fréquences fondamentales, c'est–à–dire $\omega = 2\pi\ell/n$ et $\omega' = 2\pi\ell'/n$, pour des entiers ℓ et ℓ', alors si $\omega \neq \pm\omega'$, on a que

$$E\mathcal{X}_n(\omega) \otimes \mathcal{X}_n(\omega') = \frac{1}{n}\sum_{s,t=1}^{n} E[X_s \otimes X_t]e^{-\mathrm{i}(s\omega - t\omega')} = E[X_1 \otimes X_1]\frac{1}{n}\sum_{t=1}^{n} e^{-\mathrm{i}t(\omega-\omega')} = 0,$$

et de manière similaire, $E\mathcal{X}_n(\omega) \otimes \overline{\mathcal{X}_n(\omega')} = E[X_1 \otimes X_1]\frac{1}{n}\sum_{t=1}^{n} e^{-\mathrm{i}t(\omega+\omega')} = 0$. Autrement dit, les variables complexes $\mathcal{X}_n(\omega)$ et $\mathcal{X}_n(\omega')$ sont non–corrélées, et étant par ailleurs conjointement Gaussiennes, elles sont donc indépendantes. On sait qu'en dimension finie, ces propriétés sont valides asymptotiquement dans le cas non–Gaussien et non–i.i.d. Le premier résultat de ce type pour les series fonctionnelles est dû à Panaretos et Tavakoli (2013a), or ces derniers imposaient des conditions très restrictives, basées sur les cumulants à tout les ordres et impliquant en particulier l'existence d'une infinité de moments. Nous avons suivit une approche complètement différente, développée par Peligrad et Wu (2010) pour les séries temporelles univariées. Celle-ci à l'avantage de s'appliquer à une très large classe de processus, à savoir les processus *ergodiques purement non-déterministes.* Plus rigoureusement cette dernière propriété signifie que

$$E[X_0 | \mathcal{G}_{-\infty}] \;=\; 0 \quad a.s., \tag{0.3}$$

où $\mathcal{G}_t = \sigma(X_t, X_{t-1}, \ldots)$ et $\mathcal{G}_{-\infty} = \bigcap_{t\geq 0} \mathcal{G}_{-t}$. En particulier, cette méthode permet d'étendre l'analyse spectrale à des séries temporelles fortement dépendantes. Notamment la sommabilité des covariances n'est plus nécessaire, on verra ainsi qu'il est possible de définir l'opérateur de densité spectrale de manière plus générale qu'avec (0.2). Nous avons pu généraliser leur résultat aux séries fonctionnelle, sans restrictions. Plus précisément, pour toute série temporelle fonctionnelle ergodique et purement non–déterministe $(X_t)_{t\in\mathbb{Z}}$, telle que $E\|X_0\|^2 < \infty$ et

$EX_0 = 0$, il existe *pour presque toute* fréquence $\omega \in [0, \pi[$, un opérateur $\mathcal{F}_\omega$, tel que

$$\mathcal{X}_n(\omega) \xrightarrow{d} \mathcal{CN}_H\left(0, \mathcal{F}_\omega, 0\right).$$

Comparé à Panaretos et Tavakoli (2013a), on remarque que ce résultat ne nécessite que des moments d'ordre deux. On a également une expression analogue à (0.2), mais pour la sommation de Cesàro et la convergence faible des opérateurs:

$$\mathrm{Var}(\mathcal{X}_n(\omega)) = \sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) C_h e^{-\mathrm{i}h\omega} \xrightarrow[n\to\infty]{\mathrm{w}} \mathcal{F}_\omega$$

On obtient également la convergence conjointe de $(\mathcal{X}_n(\omega_\ell))_{\ell=1,\dots,q}$ pour un nombre fini de fréquences, et donc l'indépendance asymptotique évoquée plus haut.

La preuve de Peligrad et Wu (2010) repose essentiellement sur une approximation de $\mathcal{X}_n(\omega)$ par des martingales. Intuitivement, l'idée est de change l'ordre de sommation comme suit. On définit d'abord les opérateurs de projection $\mathcal{P}_t := E[\,\cdot\,|\mathcal{G}_t] - E[\,\cdot\,|\mathcal{G}_{t-1}]$. Il vient $X_t = \mathcal{P}_t X_t + \cdots + \mathcal{P}_1 X_t + E[X_t|\mathcal{G}_0]$, puis

$$\begin{pmatrix} X_1 e^{-\mathrm{i}\omega} \\ X_2 e^{-\mathrm{i}2\omega} \\ \vdots \\ X_n e^{-\mathrm{i}n\omega} \end{pmatrix} = \begin{pmatrix} & & & \mathcal{P}_1(X_1)e^{-\mathrm{i}\omega} + & E[X_1|\mathcal{G}_0]e^{-\mathrm{i}\omega} \\ & & \mathcal{P}_2(X_2)e^{-\mathrm{i}2\omega} \quad + & \mathcal{P}_1(X_2)e^{-\mathrm{i}2\omega} + & E[X_2|\mathcal{G}_0]e^{-\mathrm{i}2\omega} \\ & \iddots & \vdots & \vdots & \vdots \\ \mathcal{P}_n(X_n)e^{-\mathrm{i}n\omega} & \cdots & \mathcal{P}_2(X_n)e^{-\mathrm{i}n\omega} + & \mathcal{P}_1(X_n)e^{-\mathrm{i}n\theta} + & E[X_n|\mathcal{G}_0]e^{-\mathrm{i}n\omega} \end{pmatrix}$$

Dans leur preuve, ils utilisent le Théorème de Carleson et une inégalité maximale de Hunt et Young pour montrer que chaque colonne ci-dessus converge *presque–sûrement* et dans $L^2(\Omega)$ vers $Z_n(\omega)e^{-\mathrm{i}n\theta}, \dots, Z_1(\omega)e^{-\mathrm{i}\omega}$. En d'autres termes, on somme en colonne plutôt qu'en ligne. Par construction $(Z_k(\omega))_{k\geq 1}$ est une différence de martingale, et la propriété (0.3) permet de montrer que $\mathcal{X}_n(\omega) = n^{-1/2}\sum_{k=1}^n Z_n(\theta)e^{-\mathrm{i}k\theta} + o_{L^2}(1)$. Dans notre situation, c'est–à–dire en fonctionnel, il est juste possible de montrer que les colonnes converges dans un sens plus faibles, à savoir dans l'espace $L^2([0,\pi[, L^2(\Omega))$. C'est tout de même suffisant pour définir *presque sûrement* l'opérateur de densité spectrale par $\mathcal{F}_\omega := \mathrm{Var}(Z_1(\omega))$. Puis, sous réserve que $\mathcal{X}_n(\omega)$ soit uniformément tendue, on sait qu'il existe une sous–suite qui converge en loi, et l'unicité de cette limite peut se déduire du résultat univarié via des projections. Pour conclure, on doit donc prouver que $\mathcal{X}_n(\omega)$ est uniformément tendue, on procède comme suit. Soit $\varepsilon > 0$, on considère des suites $0 < \ell_k \nearrow \infty$ et $0 < N_k \nearrow \infty$. Puis, on définit

$$K = \bigcap_{k=1}^{\infty} \Big\{ x \in H : \sum_{j>N_k} |\langle v_j, x\rangle|^2 \leq \ell_k^{-1} \Big\},$$

qui est un compact de H. Enfin, on a que

$$P\left(\mathcal{X}_n(\omega) \in K\right) \geq 1 - \sum_{k=1}^{\infty} \ell_k \sum_{j>N_k} E|\langle \mathcal{X}_n(\omega), v_j\rangle|^2 \tag{0.4}$$

Ensuite, on peut montrer que

$$\sup_n \sum_{j>m} E|\langle \mathcal{X}_n(\omega), v_j\rangle|^2 = \sup_n \sum_{j>m} \sum_{|h|<n} \Big(1 - |h|/n\Big) \langle C_h(v_j), v_j\rangle e^{-\mathrm{i}h\omega} \xrightarrow[m\to\infty]{} 0,$$

en utilisant le faite que $E\|\mathcal{X}_n(\omega)\|^2 \to \mathrm{Tr}(\mathcal{F}_\omega)$. Finalement, on aurait pu choisir $(\ell_k)_{k\geq 1}$ et $(N_k)_{k\geq 1}$ de telle sorte que le dernier terme dans (0.4) soit plus grand que $1-\varepsilon$.

Même si ce résultat est très satisfaisant quant à sa généralité, il s'agit d'une résultat *presque–sûr*, donc en pratique cela ne nous dit pas précisément pour quelle fréquence $\omega \in [0, \pi[$ c'est vrai. Nous avons donc travaillé sur un autre résultat similaire, mais qui lui, est valide pour toute fréquence ω, qui vérifie les hypothèses suivantes :

(A1) $M_n(\omega) := \sum_{t=0}^n \mathcal{P}_0(X_t)e^{-\mathrm{i}t\omega}$ est une suite de Cauchy dans $L^2_H(\Omega)$;

(A2) $E\Big\|E[\mathcal{X}_n(\omega)|\mathcal{G}_0]\Big\|^2 \underset{n\to\infty}{\longrightarrow} 0$.

Enfin, nous avons montré que ces hypothèses sont satisfaites dans de nombreux cas, par exemple, pour les processus linéaires, les suites L^p–m–approximables. En particulier, on peut en déduire un théorème central limite classique, pour $\omega = 0$. Enfin, ce résultat a de nombreuses applications possibles, par example, pour tester la présences de composantes périodiques comme dans Hörmann et al. (2017), pour le bootstrap, voir Paparoditis (2016) ou pour tester la stationnarité, voir Bagchi et al. (2018).

Un test de périodicité uniforme

Ensuite, nous avons étudié le comportement asymptotique du maximum de la norme de la transformée de Fourier discrète d'une série temporelle fonctionnelle, prit sur l'ensemble des fréquences. La motivation initiale de ce projet est la suivante, supposons qu'on observe

$$X_t = \mu + m_t + \varepsilon_t,$$

où $\mu \in H$, $m : \mathbb{Z} \to H$ est une fonction déterministe d–périodique et $(\varepsilon_t)_{t\in\mathbb{Z}}$ est une suite i.i.d. dans H. Pour détecter des composantes périodiques m_t d'une fréquence connue $\omega = 2\pi/d$, Hörmann et al. (2017) ont proposé un test basé sur la norme de la transformée de Fourier discrète $\mathcal{X}_n(\omega) = n^{-1/2}\sum_{t=1}^n X_t e^{-\mathrm{i}t\omega}$ de $(X_t)_{t\in\mathbb{Z}}$. Dans le but d'adapter ce test au cas où la fréquence n'est pas supposée connue d'avance, il est naturel de considérer la statistique suivante $M_n = \max_{j=1,\dots,q}\|\mathcal{X}_n(\omega_j)\|^2$, et d'étudier son comportement asymptotique sous $\mathcal{H}_0$, c'est–à–dire lorsque $(X_t)_{t\in\mathbb{Z}}$ est i.i.d. Dans le cas univarié, le résultat de Davis et Mikosch (1999) assure que M_n est dans le domaine d'attraction de la loi de Gumbel. Pour comparer avec la partie précédente, il s'agit également d'étudier la distribution conjointe de $(\mathcal{X}_n(\omega_\ell))_{\ell=1,\dots,q}$, mais lorsque q tend vers l'infini. Pour un entier $p \geq 1$, on définit la statistique tronquée $M_n^p = \max_{j=1,\dots,q}\|\mathcal{X}_n^p(\omega_j)\|^2$, où $\mathcal{X}_n^p(\omega)$ est une approximation de $\mathcal{X}_n(\omega)$ obtenu avec la représentation de *Karhunen–Loève*. Nous avons montré successivement que les trois statistiques suivantes M_n^p, $M_n^{p_n}$ (où p_n est une suite d'entiers qui tend vers l'infini) et M_n sont dans le domaine d'attraction de la loi de Gumbel. Le premier résultat généralise en multivarié le résultat de Davis et Mikosch (1999), le second précise à quelle vitesse on peut faire tendre la dimension $p = p_n$ vers l'infini, tandis que le troisième, qui nécessite des contraintes supplémentaires sur la loi de X_t, fournit une véritable généralisation fonctionnelle.

Pour obtenir ce dernier résultat, on procède comme suit. Soient $(a_n)_{n\geq 1}$ et $(b_n)_{n\geq 1}$ des suites de réels, on a que

$$a_n^{-1}\left(M_n^X - b_n\right) = a_n^{-1}\left(M_n^X - M_n^{X,p}\right) \; + \; a_n^{-1}\left(M_n^{X,p} - b_n\right). \tag{0.5}$$

Le premier terme converges vers zéro en probabilité si p croît suffisamment vite. Pour le second terme, on calcule

$$\left|P\left(a_n^{-1}(M_n^p - b_n) \leq x\right) - e^{-e^{-x}}\right| \leq \; \rho_{n,p} \; + \left|P\left(a_n^{-1}(\widetilde{M}_n^p - b_n) \leq x\right) - e^{-e^{-x}}\right|$$

où $\widetilde{M}_n^p$ est défini de manière analogue à M_n^p, mais à partir d'un échantillon Gaussien $Y_1, \ldots, Y_n$ dans H de même opérateur de variance C que X_t, et $\rho_{n,p} = \sup_{x\in\mathbb{R}} |P(M_n^p \leq x) - P(\widetilde{M}_n^p \leq x)|$. Pour prouver que le second terme converge, on remarque que pour un entier $p \geq 1$ fixé, $\|\mathcal{X}_\omega^p\|^2$ suit une loi hypo–exponentielle, or on sait déjà que cette loi appartient au domaine d'attraction de la loi de Gumbel. Puis, nous avons montré que ceci tient toujours lorsque $p = o(n)$, $a_n = \lambda_1$ et $b_n = \lambda_1 \log(n/2) - \lambda_1 \sum_{j=2}^{\infty} \log(1 - \lambda_j/\lambda_1)$, où $(\lambda_j)_{j\geq 1}$ sont les valeurs propres C ordonnées de manière décroissantes.

Pour borner $\rho_{n,p}$, on a utilisé un résultat de Chernozhukov et al. (2017). Il s'agit d'une approximation Gaussienne de $P(n^{-1/2}\sum_{i=1}^n \xi_i \in A)$, où les $\xi_1, \ldots, \xi_n$ sont des vecteurs aléatoires i.i.d. et A parcours la classe des ensembles *s–partiellement convexes* de $\mathbb{R}^d$, c'est–à–dire des intersections d'ensembles dont les indicatrices dépendent au plus de s variables. Ce résultat repose essentiellement sur une inégalité d'anti–concentration due à Chernozhukov et al. (2017). Dans notre situation, on obtient

$$\rho_{n,p} \; \leq \; \kappa \cdot \frac{p^3 \log(n)}{\lambda_p^{1/2} n^{1/6}}, \qquad \text{pour une certaine constante } \; \kappa > 0.$$

Dans le cas fonctionnel, on doit faire tendre p vers l'infini, et on a donc un terme supplémentaire au dénominateur λ_p, qui tend vers zéro. En utilisant cette borne, on a montré qu'il existe une suite d'entiers $p = p_n$, telle que les deux termes dans (0.5) convergent, ceci sous des hypothèses supplémentaires sur les moments de X_t. Plus précisément, le premier converge vers zéro en probabilité et le second en loi vers une variable de Gumbel. On peut conclure par le lemme de Slutzky.

Finalement, en dimension finie, nous avons également considéré le cas non–i.i.d. à savoir pour des processus linéaires. Puis nous avons montré comment étendre le test de Hörmann et al. (2017) lorsque la fréquence est inconnue.

Le modèle GARCH fonctionnel

Les processus GARCH univariés et multivariés modélisent la dynamique de la *volatilité*, c'est–à–dire de la variance conditionnelle dans les séries financières, ce qui permet notamment de mesurer le risque. Le modèle a été initiallement proposé par Engle (1982), il en existe aujourd'hui de très nombreuses variantes et extensions, pour une présentation générale on peut consulter Francq et Zakoïan (2011). A ce jour, les seules généralisations fonctionnelles sont Hörmann et al. (2013) pour le modèle ARCH(1) et Aue et al. (2016) pour le modèle

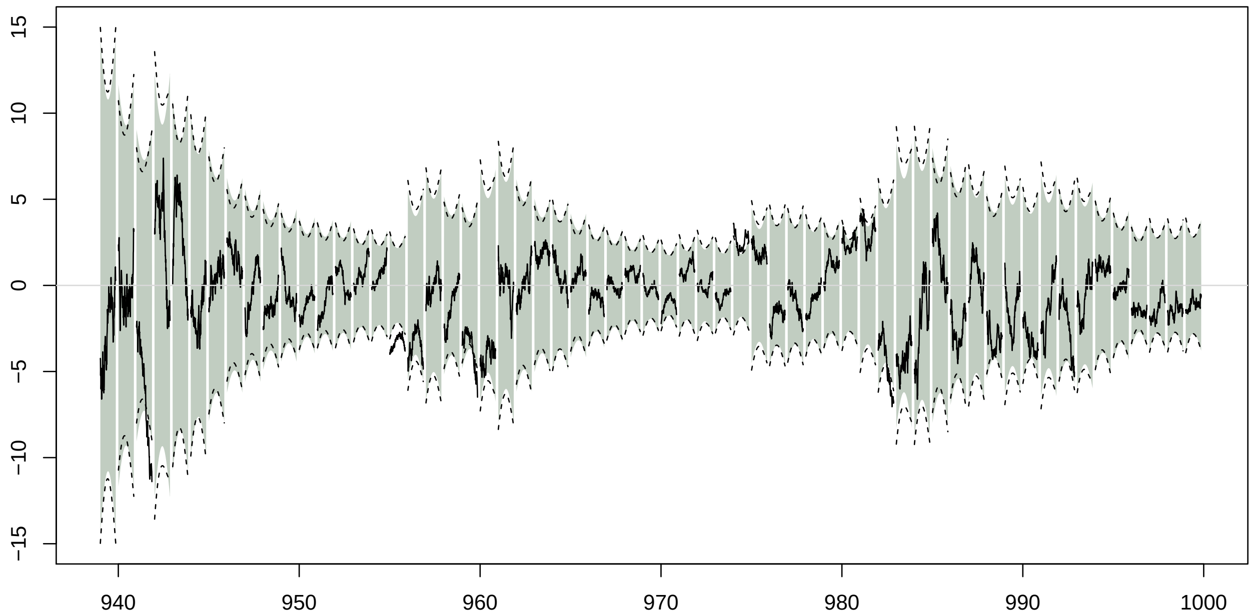

Figure 4: Le processus simulé y_t est représenté en noir, la région $\{[-2\sigma_t(u), 2\sigma_t(u)] \colon u \in [0,1]\}$ en gris et $\pm 2\widetilde{\sigma}_t(\hat{\theta})(u)$ en pointillé.

GARCH(1,1). Notre modèle est similaire à ce dernier, puis nous l'avons étendu aux processus GARCH(p,q) pour p ou $q > 1$, et considéré une approche différente pour l'estimation. Soit $(\eta_t)_{t\in\mathbb{Z}}$ une suite i.i.d. dans $L^2[0,1]$. Un processus GARCH(1,1) fonctionnel $(y_t)_{t\in\mathbb{Z}}$ est définit comme une solution stationnaire du système suivant

$$y_t(u) = \sigma_t(u)\eta_t(u), \tag{0.6}$$

$$\sigma_t^2(u) = \delta(u) + \int_0^1 K_{\boldsymbol{\alpha}}(u,v) y_{t-1}^2(v)dv + \int_0^1 K_{\boldsymbol{\beta}}(u,v)\sigma_{t-1}^2(v)dv, \tag{0.7}$$

où $\delta(u) > 0$, et $K_{\boldsymbol{\alpha}}(u,v)$, $K_{\boldsymbol{\beta}}(u,v) \geq 0$, pour tout u, $v \in [0,1]$. On dit que $\boldsymbol{\alpha}$ et $\boldsymbol{\beta}$ sont des opérateurs positifs à noyaux. Les *courbes de volatilité* peuvent s'interpréter ponctuellement car

$$P\Big(|y_t(u)| < \sigma_t(u) \cdot Q_{1-\alpha/2}^{\eta(u)} \,\Big|\, y_s,\, s < t\Big) = 1 - \alpha, \quad \forall u \in [0,1],$$

où Q_α^X est le quantile d'ordre α de la variable X.

Comme dans le cas univarié, le coefficient $\boldsymbol{\alpha}$, qui est un opérateur, est responsable de la sensibilité aux chocs de la volatilité puisque il induit une dépendance entre la courbe de volatilité σ_t^2 et le carré de la courbe de la veille y_{t-1}^2. Tandis que l'opérateur $\boldsymbol{\beta}$, est à l'origine de la persistance de la volatilité puisqu'il induit l'auto–régression de celle-ci. On peut observer ces phénomènes sur la Figure 4 où on a représenté des simulations d'un processus (0.6)–(0.7). Le processus d'innovations est Gaussien avec la condition d'identifiabilité $E[\eta^2(u)] = 1$ pour tout $u \in [0,1]$, ce qui implique que $Q_{1-\alpha/2}^{\eta(u)} \approx 2$ pour le niveau $\alpha = 0.05$. Nous avons ensuite montré qu'une condition suffisante d'existence d'une solution stationnaire, non-anticipative et ergodique $(y_t)_{t\in\mathbb{Z}}$ à (0.6)–(0.7) est que $\gamma < 0$, où

$$\gamma = \lim_{t\to\infty} \frac{1}{t} \log \|\Psi_t \dots \Psi_0\| \quad \text{et} \quad \Psi_t : \begin{cases} L^2[0,1] & \to \ L^2[0,1] \\ x & \mapsto \ \boldsymbol{\alpha}(x\cdot\eta_t^2) + \boldsymbol{\beta}(x). \end{cases}$$

Cette condition est plus générale que celle de Aue et al. (2016), à savoir que $E\log\|\Psi_0\|_{\mathcal{S}} < 0$, où $\|\cdot\|_{\mathcal{S}}$ désigne la norme *Hilbert–Schmidt*. On appelle γ l'exposant de Lyapounov associée à la suite d'opérateurs $(\Psi_t)_{t\in\mathbb{Z}}$. En dimension finie la définition ci-dessus est indépendante de la norme matricielle utilisée. Ceci n'est pas forcément vrai en dimension infinie, on utilise ici la norme d'opérateur $\|\boldsymbol{\alpha}\| = \sup_{\|x\|\le 1}\|\boldsymbol{\alpha}(x)\|$ dans $L^2[0,1]$.

Les méthodes d'estimation pour les données fonctionnelles sont généralement non–paramétriques, et essentiellement basées sur la méthode des moments, voir par exemple Bosq (2000) ou Horváth et Kokoszka (2012). La complexité du modèle GARCH rend cette approche compliquée, notamment car l'écriture explicite d'une solution, $\sigma_t^2 = \delta + \sum_{k=1}^{\infty}\Psi_t\cdots\Psi_{t-k+1}(\sigma_{t-k+1}^2)$, dépend non–linéairement des paramètres. Une possibilité est d'utiliser la représentation suivante du modèle :

$$y_t^2 - (\boldsymbol{\alpha}+\boldsymbol{\beta})(y_{t-1}^2) = \delta + \eta_t - \boldsymbol{\beta}(\eta_{t-1}), \tag{0.8}$$

où $\eta_t = y_t^2 - \sigma_t^2$ est une *différence de martingale*. Autrement dit, le processus des carrés $X_t = y_t^2 - (\mathrm{id} - \boldsymbol{\alpha} - \boldsymbol{\beta})^{-1}\delta$, est un modèle ARMA(1,1) fonctionnel. C'est la composante MA(1), corresondant ici à l'opérateur $\boldsymbol{\beta}$, qui pose problème. Nous avons donc essayé dans un premier temps, de développer une méthode d'estimation purement fonctionnelle du modèle MA(1). Les résultats obtenus dans cette direction n'ont pas été satisfaisant. De plus, les méthodes d'estimation du modèle GARCH via la représentation (0.8) comportent de nombreux inconvénients, notamment des moments trop élevés. Nous avons donc adopté une paramétrisation du modèle comme dans Aue et al. (2016). C'est–à–dire que les *paramètres fonctionnels* $(\delta, \boldsymbol{\alpha}, \boldsymbol{\beta})$ sont identifiés à un élément θ d'un compact $\Theta \subset \mathbb{R}^d$. Leur méthode reposant sur les moindres carrés, la normalité asymptotique de l'estimateur nécessite des moments d'ordre 8 ce qui n'est pas satisfaisant pour les processus GARCH. Nous avons donc plutôt essayé de développer une approche qui se rapproche autant que possible du *quasi–maximum de vraisemblance*, bien qu'il ne soit pas possible de définir véritablement la fonction de vraisemblance en dimension infinie. Pour ce faire, on considère des fonctions à valeurs positives $\varphi_1, \dots, \varphi_M$ in $L^2[0,1]$ et on définit l'estimateur $\widehat{\theta}_n$ comme le minimiseur du critère suivant

$$\widetilde{Q}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_t(\theta), \quad \tilde{\ell}_t(\theta) = \sum_{m=1}^{M}\left\{\frac{\langle y_t^2, \varphi_m\rangle}{\langle\tilde{\sigma}_t^2, \varphi_m\rangle} + \log\langle\tilde{\sigma}_t^2, \varphi_m\rangle\right\},$$

où la *volatilité empirique* $\tilde{\sigma}_t^2$ est définie récursivement via l'équation (0.7) pour des fonctions initiales y_0 et $\tilde{\sigma}_0$ dans $L^2[0,1]$. Sous l'hypothèse forte que δ, $\boldsymbol{\alpha}$ et $\boldsymbol{\beta}$ sont de rangs finis, on peut mettre en relation notre modèle avec un processus multivarié auxiliaire, qui s'avère être un *CCC–GARCH semi–fort*. Ce modèle multivarié est basé sur l'hypothèse que les corrélations conditionnelles du processus sont constantes, voir Francq et Zakoïan (2011) pour plus de détails, et *semi–fort* signifie que ses innovations ne sont pas i.i.d. Le cas semi–fort n'ayant encore jamais abordé pour les CCC–GARCH, nous avons étudié la convergence et la normalité asymptotique de l'estimateur de quasi–maximum de vraisemblance dans ce cas. Enfin, nous avons montré qu'on pouvait déduire la convergence et la normalité asymptotique de notre estimateur $\widehat{\theta}_n$, de celle de ce processus auxiliaire.

D'un point de vue fonctionnel, l'hypothèse que δ, $\boldsymbol{\alpha}$ et $\boldsymbol{\beta}$ soient de rangs finis n'est pas très satisfaisante. Nous avons donc essayé d'adapter la méthode précédente au cas véritablement fonctionnel. Il n'est alors plus possible de rapporter $(y_t)_{t\mathbb{Z}}$ à un processus auxiliaire multivariée. Mais en supposant que Θ est un compact, par exemple de l'espace des suites de carrés sommables ℓ^2, et que les paramètres fonctionnels δ, $\boldsymbol{\alpha}$ et $\boldsymbol{\beta}$ sont en quelque sorte identifiables par une famille de fonctions positives $(\varphi_m)_{m\geq 1}$, on peut montrer que l'estimateur $\widehat{\theta}_n$ est toujours convergent. Pour fixer les idées, considérons l'exemple suivant, où les opérateurs $\boldsymbol{\alpha}$ et $\boldsymbol{\beta}$ sont considérablement simplifiés. Soit $(\psi_k)_{k\geq 1}$ une base orthonormal de $L^2[0,1]$, on suppose que la volatilité vérifie l'équation suivante

$$\sigma_t^2(u) = \exp\left(\sum_{k=1}^{\infty} d_k \psi_k(u)\right) + a\int_0^1 y_{t-1}^2(v)dv + b\int_0^1 \sigma_{t-1}^2(v)dv. \tag{0.9}$$

Ici le paramètre $\theta = (a, b, d_1, d_2, \dots) \in \mathbb{R}_+^2 \times \mathbb{R}^\infty$. D'autre part, ce modèle permet une interprétation intéressante de la fonction intercept δ, en effet, si $Ey_t^2(u) < \infty$, on a que

$$E\sigma_t^2(u) = Ey_t^2(u) = \delta(u) + \frac{a+b}{1-a-b}\int_0^1 \delta(v)dv.$$

On trouve donc que δ est proportionnelle au profil de la volatilité moyenne. On peut également calculer explicitement l'exposant de Lyapounov, $\gamma = E\log(a\int_0^1 \eta_0^2(v)dv + b)$.

Finalement, nous avons testé cette méthode sur des simulations et on a pu constater son avantage sur l'estimateur des moindres carrés de Aue et al. (2016). On l'a également appliqué sur de vraies données et on a proposé une méthode de prédiction de la *volatilité réalisée*.

Asymptotic inference in stationary functional processes

1 Introduction

Due to the leap in information technology within the last decades, recording and processing of data has seen a huge and steady upsurge. The immense data flow demands for new statistical methods which can help to extract useful information. One of the modern statistical disciplines which is devoted to these new challenges is *functional data analysis (FDA)*. As the name indicates, its paradigm is to consider observations as realizations of *random functions* $u \mapsto X(u)$. Typically it is assumed that a generic functional observation X belongs to a specific space H of *real valued functions* defined on $[0,1]$, although its domain might be any continuum.
The principal area of this dissertation is the study of *functional time series (FTS)*. A FTS is a stochastic process $(X_t)_{t\in\mathbb{Z}}$ taking its values in some *function space*. Each observation is a curve $X_t(u)$. Since the data of interest are sampled sequentially in time we need to expect dependence across curves. As in classical time series problems for real data, this requires different methods than in statistics of random samples.
In this thesis we will study some inferential problems related to functional time series. These require some novel asymptotic results, which constitute the main research contribution of this account. In this section we briefly survey some background on FDA and FTS. Moreover, we set up the mathematical framework and some basic notation. Finally, we will quickly introduce the key results of this dissertation.

1.1 Discrete vs. functional data

It is a natural mathematical approach to consider the FDA framework when the underlying data generating mechanism is a continuous time process. However, before the recent technological advances, related statistical methods had not much practical importance, as neither relevant data could be recorded and stored, nor was there any hope to computationally process the resulting ‘curves’ in a satisfactory manner. The situation has now changed. Functional data analysis has been popularized, e.g. through the book of Ramsay and Silverman (2006) and the related statistical software tools which allow to conveniently process such type of data.
In practise, however, we cannot fully observe curves, but only a discretized record of the data, on which some preprocessing is performed to get a functional representation. In Figure 1 we have respectively plotted a rough and smoothed sub–sample from a popular data set introduced by Ramsay and Silverman (2006). It consists of measurements of the hip angles of 39 children at 20 time points in a single gait cycle.
While working with raw data is in principle an option (then we are having a multivariate data set), it is often ruled out by obvious disadvantages. First, the grid $\{u_{t,\ell}, 1 \le \ell \le N_t\}$ on which we observe $X_t(u_{t,\ell})$ may vary across time t. Neither dimensions N_t, nor spacings $u_{t,\ell} - u_{t,\ell-1}$ need to match. A typical example are growth curves, where the height of children is measured on irregular points in time and at different frequencies. Second, multivariate analysis makes no use of the sequential order of the components and the underlying smoothness—typically $X_t(u_{t,\ell})$ and $X_t(u_{t,\ell'})$ will be close if $u_{t,\ell}$ and $u_{t,\ell'}$ are close. It is then advantageous to approximate raw data by some curve, i.e. we approximate raw data $X_t(u_{t,\ell})$ by some finite dimensional curve $\sum_{k=1}^{p} \mathbf{x}_{t,k}\varphi_k(u)$. The most popular choices for the basis functions φ_k are

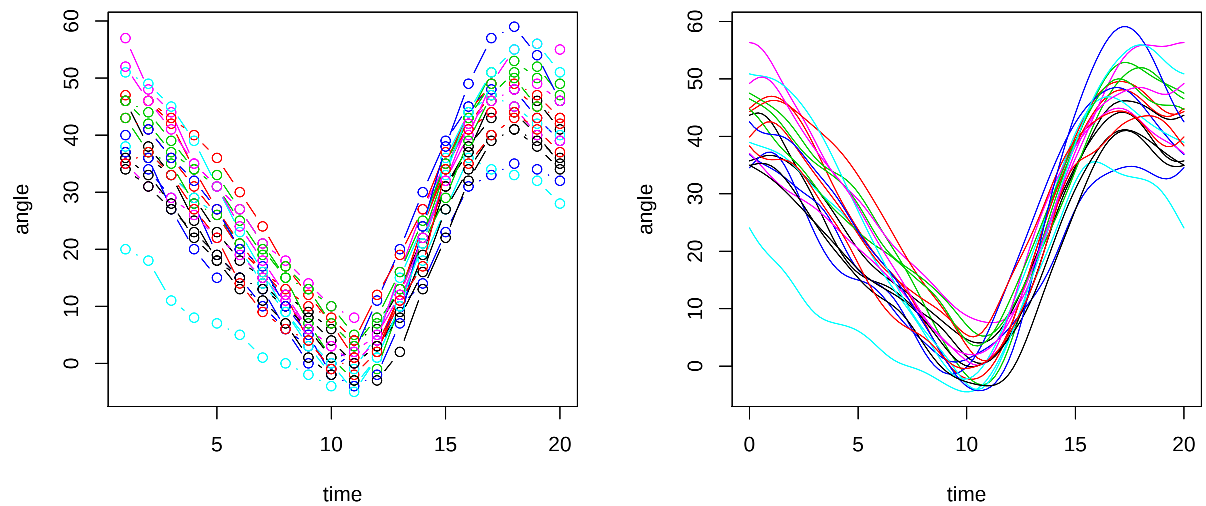

Figure 5: Hip angles of 20 children at 21 time points in a single gait cycle. Discrete observations (left panel) and smoothed sample (right panel).

Fourier basis, splines, wavelets and polynomials. A common algorithm to determine the coefficients $\mathbf{x}_{t,k}$ is by minimizing an expression of the form:

$$\sum_{\ell=1}^{N_t} \left| X_t(u_{t,\ell}) - \sum_{k=1}^{p} \mathbf{x}_{t,k}\varphi_k(u_{t,\ell}) \right|^2 + \lambda \int_0^1 \left| \sum_{k=1}^{p} \mathbf{x}_{t,k}\varphi_k''(u) \right|^2 du. \tag{1.1}$$

In other words, a penalized regression is used. The penalty in (1.1) allows to control the smoothness of the curves. The coefficient $\lambda > 0$, the so–called *roughness penalty*, has thus to be tuned in order to make a trade-off between bias and variance. We refer again to Ramsay and Silverman (2006) for basic information. Note that the resulting coefficients $\mathbf{x}_1, \ldots, \mathbf{x}_n$ form again a multivariate sample. However, a key difference with traditional multivariate analysis is that the number p is not fixed and has to be tuned.
In this thesis we bypass the problem of fitting raw data by assuming fully observed curves. This approach is quite common in FDA when the target is to derive theoretical results.

1.2 Mathematical framework

Once we have our data in fully functional form, we need to embed them in some appropriate algebraical and topological framework. As it is common in FDA, we will assume throughout this thesis that the observations belong to a *separable real Hilbert space H*. This setting, which includes in particular $L^2[0,1]$—the space of square integrable functions on $[0,1]$—is mathematically very convenient due to the nice geometrical properties of Hilbert spaces. Each such space H is endowed with a scalar product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$. The separability ensures the existence of denumerable *orthonormal basis*. The projection onto basis functions is very useful to make the link with a multivariate setting.

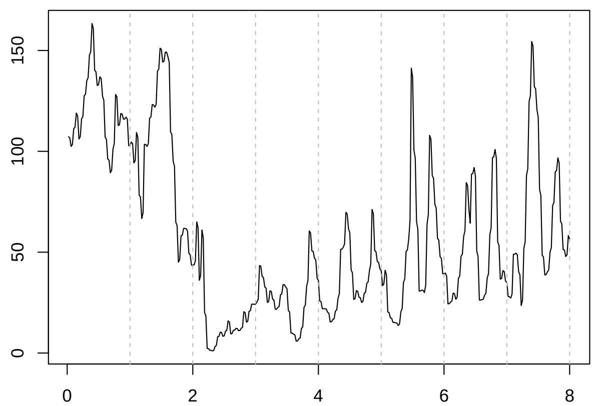

Figure 6: PM_{10} ($\mu g/m^3$) in Graz, 30–minutes resolution, 18–16/01/2010.

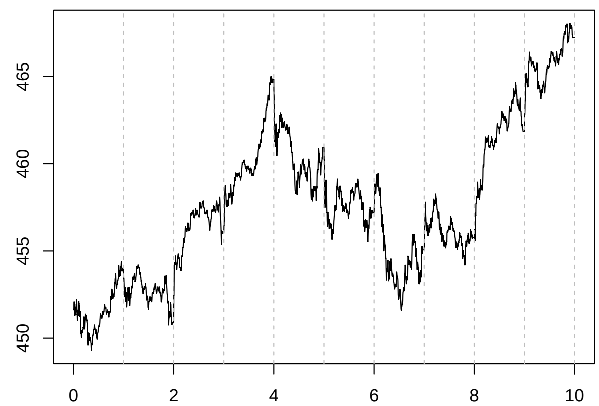

Figure 7: S&P market index in one–minute resolution, 15–29/01/1998.

Now that the space is fixed on which we will operate, we want to equip it with a few important tools. In particular, our theoretical developments will heavily make use of linear operators. Throughout this thesis we denote by $\mathcal{L}(H)$ the space of *bounded linear operators* in H, i.e., the set of continuous linear mappings $\boldsymbol{\alpha} : H \to H$. For some $\boldsymbol{\alpha} \in \mathcal{L}(H)$ we define the *operator norm* $\|\boldsymbol{\alpha}\|_{\mathcal{L}} = \sup_{\|x\|\leq 1} \|\boldsymbol{\alpha}(x)\|$. A simple way to construct a linear operator is by an outer product of two elements $x, y \in H$. Then $x \otimes y : x \mapsto x\langle \cdot, y\rangle$. We define the *adjoint operator* $\boldsymbol{\alpha}^*$, as the unique element in $\mathcal{L}(H)$ such that $\langle x, \boldsymbol{\alpha}(y)\rangle = \langle \boldsymbol{\alpha}(x), y\rangle$ for all $x, y \in H$. If, for all $x \in H \setminus \{0\}$, we have that $\langle x, \boldsymbol{\alpha}(x)\rangle \geq 0$, respectively $\langle x, \boldsymbol{\alpha}(x)\rangle > 0$, we say that $\boldsymbol{\alpha}$ is *non-negative*, respectively *non-negative definite*. We say that an operator is *compact* if the image of the unit ball has compact closure. If $\boldsymbol{\alpha}$ is compact and self-adjoint, then by the spectral theorem (see e.g. in Brezis (2010)) we know that there exists a sequence of real numbers $(\lambda_k)_{k\geq 1}$ and an orthonormal basis of H $(v_k)_{k\geq 1}$ such that

$$\boldsymbol{\alpha} \;=\; \sum_{k=1}^{\infty} \lambda_k v_k \otimes v_k. \tag{1.2}$$

Moreover, if $\boldsymbol{\alpha}$ is non-negative, then the λ_k's are non-negative. Let $(\ell_k)_{k\geq 1}$ denote the singular values of an operator $\boldsymbol{\alpha}$, i.e. the square roots of the eigenvalues of $\boldsymbol{\alpha}\boldsymbol{\alpha}^*$. We denote by $\mathcal{S}_H$ the space of *Hilbert–Schmidt* operators in H, i.e. operators whose *Hilbert–Schmidt* norm $\|\boldsymbol{\alpha}\|_{\mathcal{S}}^2 = \sum_{k=1}^{\infty} \ell_k^2$ is finite. Moreover, we let $\mathcal{T}(H)$ be the space of *trace–class* operators in H, i.e. operators whose *trace–class* norm $\|\boldsymbol{\alpha}\|_{\mathcal{T}} = \sum_{k=1}^{\infty} \ell_k$ is finite. Both $\mathcal{S}(H)$ and $\mathcal{T}(H)$ are Banach spaces with their corresponding norms. More precisely, $\mathcal{S}(H)$ is a Hilbert space whose scalar product is given by $\langle \boldsymbol{\alpha}, \boldsymbol{\beta}\rangle_{\mathcal{S}} = \sum_{k=1}^{\infty} \langle \boldsymbol{\alpha}(v_k), \boldsymbol{\beta}(v_k)\rangle$, and this sum is independent the orthonormal basis $(v_k)_{k\geq 1}$. Note that Hilbert-Schmidt operators and trace-class operators are compact. When $H = L^2[0,1]$, we say that $\boldsymbol{\alpha} \in \mathcal{L}(H)$ is a *kernel operator* since there exists a function $K_{\boldsymbol{\alpha}} : [0,1] \times [0,1] \to \mathbb{R}$ such that $\boldsymbol{\alpha}(x)(u) = \int_0^1 K_{\boldsymbol{\alpha}}(u,v)x(v)dv$. In this situation, we have that $\boldsymbol{\alpha} \in \mathcal{S}(H)$ if and only if $K_{\boldsymbol{\alpha}} \in L^2([0,1] \times [0,1])$. For more details, see.g. Bosq (2000).

Up to here, we have only described the space H on which our functional data take there realizations. As a matter of fact, the objective is to model functional data as random elements in H. For the entire thesis we assume that all our random curves are defined on a rich enough probability space $(\Omega, \mathcal{A}, P)$. Then a generic observation X is a measurable mapping $X : (\Omega, \mathcal{A}) \to (H, \mathcal{B}_H)$, where $\mathcal{B}_H$ is the Borel sigma algebra of H, i.e. the σ-algebra generated by the open sets in H. There is a well developed tool kit for probability theory on Banach and Hilbert space. Excellent reference are Ledoux and Talagrand (1991) or Bosq (2000). In particular the latter account is a key reference for our thesis, since it explores a *functional time series* oriented approach. In general, the infinite dimensional nature of H complicates some aspects when compared to the finite dimensional setup. For example, a common task for a mathematical statistician is verification of weak convergence of some statistic of interest. Depending on the problem at hand, verification of *tightness* can be quite challenging in the infinite dimensional setup. We will meet such problems later in this thesis.
Another fundamental distinction with the finite dimensional setting is that there is no equivalent to the probability density function of a random variable in H, and thus the classical *likelihood approaches*—which are fundamental for statistical inference—do not exist.

1.3 Temporal dependence

As we have noted above, we are interested here in inference for *functional time series*. Functional time series are often obtained from some continuous time process which can be naturally chopped into units of a certain length. For example, if solar radiation is recorded, then it is natural to analyse the diurnal variation and a 24h time interval is a natural unit. Another example of this type are pollution level curves which are impacted by day-to-day traffic routines. In Figure 6 we have plotted 8 curves representing the diurnal PM_{10} concentration. Raw data for this plot consist of 48 values per day. In Figure 7 we have plotted 10 daily curves from the S&P market index, here 405 values per trading day are recorded.
We will assume that the functional time series treated in this thesis are *stationary*, i.e. that $(X_{t_1}, \ldots, X_{t_k}) \overset{d}{=} (X_{t_1+h}, \ldots, X_{t_k+h})$, for all $t_1, \ldots, t_k, h \in \mathbb{Z}$ and all $k \geq 1$. However, unlike it is typical in FDA context, we do not want to assume that the resulting observations are i.i.d., but rather expect temporal dependence across days. This dependence needs to be taken into account for the statistical analysis. Thus, we will be required to update accordingly the probabilistic toolbox for dealing with such data. For example, it is important to know if we can still employ the *law of large numbers* or the *central limit theorem*—both being fundamental ingredients for statistical inference. The answer to such questions relies very much on the dependence structure which is underlying the data generating mechanism. For example, under finite second order moments, i.e. if $E\|X_0\|^2 < \infty$, the most basic dependence measures of a stationary functional time series are the *lagged auto-covariance operators*, defined as follows

$$C_h \;=\; E[(X_h - EX_0) \otimes (X_0 - EX_0)], \quad \text{for all} \quad h \in \mathbb{Z}. \tag{1.3}$$

Note that C_h are Hilbert Schmidt operators and, in particular, C_0 is a trace class, self-adjoint and non-negative operator, such that $\|C_0\|_{\mathcal{T}} = E\|X_0\|^2$. It is quite easy to derive the weak law of large numbers by routine arguments from $\|C_h\| \to 0$ for $h \to \infty$. Under *ergodicity* we can even derive the *strong law of large numbers*. To this end we recall that a stationary process $(X_t)_{t\in\mathbb{Z}}$ can be represented as a dynamical system. There exists a random element ξ in H defined on a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P})$ and a transformation $T : \widetilde{\Omega} \to \widetilde{\Omega}$ such that $\widetilde{P} \circ T^{-1} = \widetilde{P}$ and $X_t \stackrel{d}{=} \xi \circ T^t$. We say that $(X_t)_{t\in\mathbb{Z}}$ is *ergodic* if the T–invariant sets in $\widetilde{\Omega}$, i.e. the sets $A \in \widetilde{\Omega}$ such that $T^{-1}(A) = A$, have $\widetilde{P}$–measure equal to zero or one. Then the process satisfies the ergodic theorem, i.e.

$$\frac{1}{n}\sum_{t=1}^{n} \xi \circ T^t \underset{n\to\infty}{\longrightarrow} E\xi, \quad \widetilde{P}\text{–almost surely}, \tag{1.4}$$

provided that $\int_{\widetilde{\Omega}} \|\xi\| d\widetilde{P} < \infty$. Such results from ergodic theory are more commonly stated for $\mathbb{R}$–valued random processes, but it is easy to extend them to Hilbert spaces, see e.g. Parthasarathy (1967).

A special class of ergodic processes is constituted by so-called *Bernoulli shifts.* A Bernoulli shift is defined as

$$X_t = f(\ldots, \varepsilon_{t-1}, \varepsilon_t, \varepsilon_{t+1}, \varepsilon_{t+2}, \ldots), \tag{1.5}$$

where the ε_t's are i.i.d. random variables in H. A *causal* Bernoulli shift has representation $X_t = f(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$. Here $f : H^\infty \to H$ is a $\mathcal{B}_H^\infty$–$\mathcal{B}_H$ measurable function. Many important time series models have such a Bernoulli shift representation, and we may conclude that the strong law of large numbers is applicable.

The most important examples for FTS with a Bernoulli shift represenation are *functional autoregressive* models and *functional moving average* models. These are defined as

$$X_t \;=\; \Phi(X_{t-1}) + \varepsilon_t \quad \textit{and} \quad X_t \;=\; \Theta(\varepsilon_{t-1}) + \varepsilon_t, \tag{1.6}$$

respectively, where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence in H and $\Phi, \Theta \in \mathcal{L}_H$. The autoregressive model has been the focus of many publications (see e.g. Mas (2007); Kowal et al. (2016); Aue et al. (2015); Guillas et al. (2011); Kargin and Onatski (2008)), whereas the moving average has only been approached rarely (e.g. by Turbillon (2007) and Aue and Klepsch (2017a)). Clearly, the functional MA process has the form (1.5). Under mild regularity assumptions on Φ, also the functional AR process can be represented as a *functional linear process* in H, i.e. there is a sequence of operators $(\Psi_\ell)_{\ell\geq 1}$ such that

$$X_t \;=\; \sum_{\ell\geq 0} \Psi_\ell(\varepsilon_{t-\ell}). \tag{1.7}$$

Functional linear processes have been comprehensively explored in Bosq (2000). Many important asymptotic results have been developed specifically for such linear time series.

In Chapter 4 we will explore *functional GARCH*, which is an example of a non–linear functional time series model being also of the form (1.5). The functional GARCH is defined as a stationary solution $(y_t)_{t\in\mathbb{Z}}$ of the following equations

$$y_t = \sigma_t \cdot \eta_t, \quad \text{and} \quad \sigma_t^2 = \delta + \boldsymbol{\alpha}(y_{t-1}^2) + \boldsymbol{\beta}(\sigma_{t-1}^2), \tag{1.8}$$

where $\delta \in H$, $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathcal{L}_H$, and for all $x, y \in H$, we denote by $x \cdot y$ and x^2, respectively the pointwise product and square of functions, see e.g. Aue et al. (2016).
The class of Bernoulli-shifts (1.5) satisfies the law of large numbers, but is still too big to immediately guarantee a broad range of useful properties, like for example the *central limit theorem*. To overcome this difficulty we can either impose particular time series models as the ones listed above, or we further strengthen assumptions in order narrow down this class. For example, Hörmann and Kokoszka (2010) consider Bernoulli shifts which are *L^p-m-approximable*. For $p \geq 1$ we say that a process $(X_t)_{t\in\mathbb{Z}}$ is L^p-m-approximable if $E\|X_0\|^p < \infty$ and if $\sum_{m=1}^{\infty}(E\|X_0 - X_0^{(m)}\|^p)^{1/p} < \infty$, where

$$X_t^{(m)} := f\left(\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-m+1}, \varepsilon'_{t-m}, \varepsilon'_{t-m-1}, \ldots\right), \tag{1.9}$$

and $(\varepsilon'_t)_{t\in\mathbb{Z}}$ is an independent copy of $(\varepsilon_t)_{t\in\mathbb{Z}}$. For example, this property is satisfied if $(X_t)_{t\in\mathbb{Z}}$ is a functional linear process with $E\|\varepsilon_0\|^p < \infty$ and $\sum_{\ell=1}^{\infty} \ell \, \|\Psi_\ell\| < \infty$ (Hörmann and Kokoszka (2010)) or if it is a functional GARCH with $E\|y_0\|^p < \infty$ (Kokoszka et al. (2017)). Within the class of L^p-m-approximable processes, a number of inferential problems can be solved as it was shown in Hörmann and Kokoszka (2010). We will employ this property in Section 2.3.3 and Section 3.3.1.

1.4 Inference for functional time series

The most relevant inferential problems in FTS are related to mean and autocovariance operators. We have already noted above that the mean can be estimated consistently for processes of type (1.5) by the ergodic theorem. The same holds true for the autocovariance function (1.3). The variables $X_{t+h} \otimes X_t$ inherit the Bernoulli shift representation and thus consistent estimation of C_h by its empirical counterpart

$$\widehat{C}_h = n^{-1} \sum_{t=1}^{n-h} (X_{t+h} - \overline{X}_n) \otimes (X_t - \overline{X}_n) \quad \text{for } h \geq 0 \text{ and } \widehat{C}_h = \widehat{C}^*_{-h} \text{ for } h < 0, \tag{1.10}$$

is again immediate from the ergodic theorem (provided second order moments of X_t exist). It is known that convergence in the ergodic theorem can be arbitrarily slow and for statistical purposes it is thus desirable to have qualitative results about the speed of convergence. Under L^2-m-approximability it can be shown (see Hörmann and Kokoszka (2010)) that $E\|\bar{X}_n - \mu\| = O\big(1/\sqrt{n}\big)$ (here $\mu = EX_1$ and $\bar{X}_n = (X_1 + \cdots + X_n)/n$). Such a result holds also for the lagged covariance operators defined in (1.3). By Lemma D.1 Hörmann et al. (2017), if $(X_t)_{t\in\mathbb{Z}}$ is L^4–m–approximable, then $E\|\widehat{C}_h - C_h\|_{\mathcal{S}} = O\big(1/\sqrt{n}\big)$.
If we want to go a step further, we can ask if $(X_t)_{t\in\mathbb{Z}}$ satisfies a CLT. For an i.i.d. sequence in H with $E\|X_1\|^2 < \infty$ it is well known that $n^{1/2}(\bar{X}_n - \mu) \xrightarrow{d} Z$, where Z is a *Gaussian random element* in H. Hence Z is a random element such that for all $v \in H$, $\langle Z, v\rangle$ is a Gaussian random variable. See for example Bosq (2000). As we already mentioned above, the main step in the proof is the verification of tightness. In the i.i.d. setup this part is relatively easy. Assume without loss of generality that $\mu = 0$. Let $\varepsilon > 0$ and consider some sequences $0 < \ell_k \nearrow \infty$ and $0 < N_k \nearrow \infty$. We then define

$$K = \bigcap_{k=1}^{\infty} \Big\{ x \in H : \sum_{j > N_k} |\langle v_j, x\rangle|^2 \leq \ell_k^{-1} \Big\},$$

with λ_j being the eigenvalues of $C_0 = \mathrm{Var}(X_1)$ and v_j the corresponding eigenfunctions. It is readily verified that this is a compact subset of H. Then we have that

$$P\left(n^{1/2}\bar{X}_n \in K\right) \geq 1 - \sum_{k=1}^{\infty} \ell_k \sum_{j>N_k} E\Big|\langle n^{1/2}\bar{X}_n, v_j\rangle\Big|^2 = 1 - \sum_{k=1}^{\infty} \ell_k \sum_{j>N_k} \lambda_j.$$

We may always choose the sequences $(\ell_k)_{k\geq 1}$ and $(N_k)_{k\geq 1}$ in such a way that the last term is greater that $1-\varepsilon$. Here we can see that tightness is closely related to the summability of the eigenvalues λ_j of C_0. In Chapter 2 we will give a very general extension of this result to stationary time series.

One key tool in functional data analysis is the *functional principal component analysis (FPCA)*. Functional principal components are simply the eigenfunctions of C_0. The expansion of X along this basis yields the celebrated *Karhunen–Loève expansion*:

$$X \;=\; \sum_{k=1}^{\infty} \langle X, v_k\rangle v_k. \tag{1.11}$$

When we expand along the first $p \geq 1$ eigenfunctions only, we obtain a finite dimensional approximation which is optimal in the sense that

$$E\|X - \sum_{k=1}^{p} \langle X, v_k\rangle v_k\|^2 \leq E\|X - \sum_{k=1}^{p} \langle X, b_k\rangle b_k\|^2$$

for any p and any orthonormal basis $(b_k)_{k\geq 1}$. This suggests to project functional objects of interest onto the space spanned by the first p eigenfunctions and to employ multivariate tools in a first step. In a second step we let p increase with sample size. This tuning has to be done very carefully and is one of the key difficulties of functional data inference. Let us exemplify this approach in estimating Φ in functional autoregressive model (1.6). It can be readily seen that $C_1 = \Phi C_0$. This suggests $\Phi = C_1 C_0^{-1}$, however since C_0 is a compact operator, its inverse is not bounded. To avoid this problem, we can take the inverse on the finite dimensional subspace of H spanned by the first principal components $v_1, \ldots, v_p$, i.e. we define

$$\widehat{C_0^{-1}} = \sum_{k=1}^{p} \widehat{\lambda}_k^{-1} \widehat{v}_k \otimes \widehat{v}_k, \tag{1.12}$$

where $\widehat{\lambda}_k$ and $\widehat{v}_k$ are the eigenvalues and eigenfunction of the empirical covariance operator $\widehat{C}_0$ given in (1.10). Finally, we define the projection operator $\widehat{\pi}_p = \sum_{k=1}^{p} \widehat{v}_k \otimes \widehat{v}_k$ and take $\widehat{\Phi} = \pi_p \widehat{C}_1 \widehat{C_0^{-1}}$, which will converge *almost surely* to Φ in $\mathcal{L}_H$, under some technical assumptions on the eigenvalues $(\lambda_k)_{k\geq 1}$ and the sequence $p = p_n \nearrow \infty$, see e.g. Bosq (2000). Furthermore, if $H = L^2[0,1]$ and Φ is a kernel operator, then we have that

$$K_{\widehat{\Phi}}(x,y) \;=\; \frac{1}{n-1} \sum_{t=1}^{n-1} \sum_{k=1}^{p} \sum_{\ell=1}^{p} \widehat{\lambda}_k^{-1} \langle \widehat{v}_k, X_t\rangle \langle \widehat{v}_\ell, X_{t+1}\rangle \widehat{v}_k(y) \widehat{v}_\ell(x).$$

More details about inference for FTS can be found e.g. in Horváth and Kokoszka (2012).

1.5 FTS in the frequency domain

Like in classical univariate and multivariate time series analysis, it is natural to also investigate functional times series in the frequency domain. A variety of problems can be solved more easily in the frequency rather than in the time domain. In frequency domain analysis, rather than working on the sample $X_1, \dots, X_n$, being indexed in time, we focus on its *discrete Fourier transform (DFT)* defined as

$$\mathcal{X}_n(\omega) = n^{-1/2} \sum_{t=1}^{n} X_t e^{-\mathrm{i}t\omega}, \quad \omega \in (-\pi, \pi], \tag{1.13}$$

which is indexed by frequencies. Then, the knowledge of $\mathcal{X}_n(\omega)$ at all *fundamental frequencies* $\omega = 2\pi\ell/n$ for $\ell = -\lfloor (n-1)/2 \rfloor, \dots, \lfloor n/2 \rfloor$ is equivalent to that of $(X_1, \dots, X_n)$, in the sense that one can be computed from the other. Similarly, the counterpart to the lagged covariance operators is the *spectral density operator*, defined by

$$\mathcal{F}_\omega := \sum_{h \in \mathbb{Z}} C_h e^{-\mathrm{i}h\omega}, \quad \omega \in (-\pi, \pi]. \tag{1.14}$$

If the above series converges, it can be shown that $\mathcal{F}_\omega$ is the asymptotic variance of $\mathcal{X}_n(\omega)$. The spectral density operator $\mathcal{F}_\omega$ is in turn equivalent to the auto–covariance functions. Among the numerous applications of the spectral density operator, let us mention that the representation (1.11), which only relies on the covariance at lag $h = 0$, and thus does not take the dependence into account, has a *dynamic* equivalent which relies on the full knowledge of $\mathcal{F}_\omega$. See Hörmann et al. (2015a) and Panaretos and Tavakoli (2013b).
A typical estimator of it is

$$\widehat{\mathcal{F}}_\omega = \sum_{|h| < \ell_n} \left(1 - \frac{|h|}{\ell_n}\right) \widehat{C}_h e^{-it\omega}. \tag{1.15}$$

Under technical assumptions, including L^4–m–approximability, Hörmann et al. (2015a) show that

$$\sup_\omega E \left\| \widehat{\mathcal{F}}_\omega - \mathcal{F}_\omega \right\|_{\mathcal{S}} \underset{n\to\infty}{\longrightarrow} 0. \tag{1.16}$$

To study $\mathcal{X}_n(\omega)$ we need to work on complex Hilbert spaces. Let us assume that our functional observation belong to a separable real Hilbert space H_0 and define the complex version of it as $H = H_0 \oplus \mathrm{i}H_0$. The covariance structure of a random element Z in H is characterized by *two* operators: its *complex covariance operator* $\Gamma = E[Z \otimes Z]$ and its *relation operator* $C = E[Z \otimes \overline{Z}]$, where $\overline{Z}$ denotes the complex conjugate of Z. Two centred random elements Z_1 and Z_2 in H are uncorrelated if both $E[Z_1 \otimes Z_2]$ and $E[Z_1 \otimes \overline{Z_2}]$ are zero. For example, the Gaussian distribution $\mathcal{N}_H(\mu, \Gamma, C)$ in H is characterized by $\mu \in H$, Γ and $C \in \mathcal{L}_H$. The particular case where $\mu = 0$ and $C = 0$ corresponds to the so-called *circular–symmetric* case. (In Appendix 5.1 we provide a more detailed presentation of complex normal random elements.) Suppose that $X_1, \dots, X_n$ are i.i.d. $\mathcal{N}_H(0, \Gamma, 0)$. Then $\mathcal{X}_n(\omega)$ is distributed as $\mathcal{N}_H(0, \Gamma, 0)$, for all $\omega \in [0, \pi]$. Moreover, consider two fundamental frequencies i.e. $\omega = 2\pi\ell/n$ and $\omega' = 2\pi\ell'/n$, for some integers ℓ and ℓ', then if $\omega \neq \pm\omega'$, we have that

$$E\mathcal{X}_n(\omega) \otimes \mathcal{X}_n(\omega') = \frac{1}{n} \sum_{s,t=1}^{n} E[X_s \otimes X_t] e^{-\mathrm{i}(s\omega - t\omega')} = E[X_1 \otimes X_1] \frac{1}{n} \sum_{t=1}^{n} e^{-\mathrm{i}t(\omega - \omega')} = 0,$$

and similarly, $E\mathcal{X}_n(\omega) \otimes \overline{\mathcal{X}_n(\omega')} = E[X_1 \otimes X_1]\frac{1}{n}\sum_{t=1}^n e^{-\mathrm{i}t(\omega+\omega')} = 0$. In other words, for all distinct fundamental frequencies the discrete Fourier transforms are jointly Gaussian and uncorrelated, thus, independent. We will see in Chapter 2 that this feature is asymptotically valid for a broad class of stationary processes in H, namely for the class of *purely non-deterministic* processes i.e. if

$$E[X_0|\mathcal{G}_{-\infty}] \;=\; 0 \quad a.s., \tag{1.17}$$

where $\mathcal{G}_t = \sigma(X_t, X_{t-1}, \ldots)$ and $\mathcal{G}_{-\infty} = \bigcap_{t\geq 0} \mathcal{G}_{-t}$.

1.6 Overview of the results

The infinite dimensional aspect of the topic was leading throughout the work to some mathematical challenges, commonly encountered in FDA. Let us highlight three key difficulties: (i) Verification of tightness of a sequence of random functions. (ii) Letting the dimension $p = p_n$ after spectral truncation tend to infinity. (iii) Dealing with the non-existence of likelihood function in infinite dimension. Problem (i) was mainly encountered in Chapter 2 and is closely related to the summability of the eigenvalues of the spectral density operator. Problem (ii) is somehow omnipresent, but is particularly relevant in Chapter 3. The last one, mainly occurs in estimation problems such as in Chapter 4. Our chapters are based on three scientific papers. *On the CLT for discrete Fourier transforms of functional time series*, with co–author Siegfried Hörmann, is accepted in Journal of Multivariate Analysis. *A uniform test of periodicity for functional time series* with co–authors Vaidotas Characiejus and Siegfried Hörmann, will be submitted soon. Whereas *Functional GARCH models: the quasi-likelihood approach and its applications* with co–authors Christian Francq, Siegfried Hörmann and Jean–Michel Zakoïan, is currently in the reviewing process.

Asymptotic normality of the discrete Fourier transform

In Chapter 2, we derive the asymptotic behaviour of the discrete Fourier transform $\mathcal{X}_n(\omega)$ under minimal assumptions. Note that this asymptotic behaviour is useful in many practical problems, e.g. in the detection of periodic patterns, see e.g. Hörmann et al. (2017), for bootstrapping of FTS, see e.g. Paparoditis (2016), or for testing stationarity of a FTS, see e.g. Bagchi et al. (2018). In particular, our results improve upon Panaretos and Tavakoli (2013a) who work under cumulant conditions, which require in particular infinitely many moments. We follow a completely different approach, based on the results of Peligrad and Wu (2010). We manage to generalize their univarite results to FTS. In particular, we show in Theorem 1 that for every ergodic and and purely non–deterministic and stationary processes $(X_t)_{t\in\mathbb{Z}}$ (see (1.17)) in a separable complex Hilbert space H, with $E\|X_0\|^2 < \infty$ and $EX_0 = 0$, there exists an operator $\mathcal{F}_\omega$ such that

$$\mathcal{X}_n(\omega) \xrightarrow{d} \mathcal{CN}_H\left(0, \mathcal{F}_\omega, 0\right),$$

for almost all ω in $(-\pi, \pi]$. The result does not require more than two moments, nor the summability of the covariance operators. Hence, the definition (1.14) for $\mathcal{F}_\omega$ cannot be directly used. Rather we have that

$$\mathrm{Var}(\mathcal{X}_n(\omega)) = \sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) C_h e^{-\mathrm{i}t\omega} \xrightarrow[n\to\infty]{\mathrm{w}} \mathcal{F}_\omega, \tag{1.18}$$

i.e. $\mathcal{F}_\omega$ is the limit of a Cèsaro mean for the so–called *weak operator* convergence.
We consider as well the joint convergence of $(\mathcal{X}_n(\omega_\ell))_{\ell=1,\dots,q}$ for a finite number of frequencies, and show asymptotic independence for distinct frequencies. The main trick in the proof, due to an idea in Peligrad and Wu (2010), is based on a change in the order of summation. To this end, we define projection operators $\mathcal{P}_t := E[\,\cdot\,|\mathcal{G}_t] - E[\,\cdot\,|\mathcal{G}_{t-1}]$ and note that $X_t = \mathcal{P}_t X_t + \cdots + \mathcal{P}_1 X_t + E[X_t|\mathcal{G}_0]$, and thus that

$$\begin{pmatrix} X_1 e^{-\mathrm{i}\omega} \\ X_2 e^{-\mathrm{i}2\omega} \\ \vdots \\ X_n e^{-\mathrm{i}n\omega} \end{pmatrix} = \begin{pmatrix} & & & \mathcal{P}_1(X_1)e^{-\mathrm{i}\omega} + & E[X_1|\mathcal{G}_0]e^{-\mathrm{i}\omega} \\ & & \mathcal{P}_2(X_2)e^{-\mathrm{i}2\omega} \quad + & \mathcal{P}_1(X_2)e^{-\mathrm{i}2\omega} + & E[X_2|\mathcal{G}_0]e^{-\mathrm{i}2\omega} \\ & \iddots & \vdots & \vdots & \vdots \\ \mathcal{P}_n(X_n)e^{-\mathrm{i}n\omega} & \cdots & \mathcal{P}_2(X_n)e^{-\mathrm{i}n\omega} + & \mathcal{P}_1(X_n)e^{-\mathrm{i}n\theta} + & E[X_n|\mathcal{G}_0]e^{-\mathrm{i}n\omega} \end{pmatrix}$$

We now sum over the columns, rather than rows, on the right hand side and denote the column sums by $Z_{k,n}(\omega)e^{-\mathrm{i}k\omega}$. Roughly speaking we show that $Z_{k,n}(\omega)$ converge in some appropriate sense to a $Z_k(\omega)e$. By construction $(Z_k(\omega))_{k\geq 1}$ is a stationary martingale difference sequence. Using the purely non–deterministicness assumption, one can show that $\mathcal{X}_n(\omega) = n^{-1/2}\sum_{k=1}^n Z_k(\theta)e^{-\mathrm{i}k\theta} + o_{L^2}(1)$. Then, define $\mathcal{F}_\omega := \mathrm{Var}(Z_1(\omega))$, which will be shown to coincides with the limit in (1.18). Provided that $\mathcal{X}_n(\omega)$ is tight in H, we deduce the uniqueness of the weak limit by reducing to their scalar result through projections. Finally we prove the tightness of $\mathcal{X}_n(\omega)$, which mainly relies on the fact that $E\|\mathcal{X}_n(\omega)\|^2 \to \|\mathcal{F}_\omega\|_{\mathcal{T}}$, when n tends to infinity.

This result is very sharp and is based on almost no assumptions. However, it is not very practical because we don't know for which frequency ω it does not hold. (Only if we pick one at random we can be sure that the CLT holds.) We thus provide a second result under slightly less general assumptions, Theorem 2, and which is valid for any given ω. We also show that this new assumption is still general enough, in order to be applicable for most of functional times series appearing in the literature (L^p-m-approximable processes). In particular we deduce from Theorem 2 the regular CLT ($\omega = 0$) under new and very mild assumptions.

A uniform periodicity test

In Chapter 3, we investigate the asymptotic behaviour of the maximum, over all fundamental frequencies, of the norm of the DFT. Let us first present a statistical problem that has motivated our work. Consider a functional time series of the form

$$X_t = \mu + s_t + \varepsilon_t,$$

where $\mu \in H$, $s_t\colon \mathbb{Z} \to H$ is a d–periodic deterministic sequence of functions in H, and where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a zero mean functional time series in H. If the period d is known, then Hörmann et al. (2017) have developed some statistical tests for $\mathcal{H}_0 : \|s_t\| = 0$ for all t vs. $\mathcal{H}_1\colon \exists t$ such

that $\|s_t\| \neq 0$. In some situation the period d may be unknown. In this case, a statistical test needs to be based on all potential periods. In particular, it is natural to consider the statistic $M_n = \max_{j=1,\ldots,q} \|\mathcal{X}_n(\omega_j)\|^2$ with $\omega_j = 2\pi j/n$, for $j = 1, \ldots, q = \lfloor (n-1)/2 \rfloor$, and investigate its behaviour under the null, i.e. when $(X_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence in H. This problem is delicate. Note that the results of Chapter 2 are not applicable here, because M_n^X depends on an increasing number of frequencies. Here we somehow need to understand the joint behaviour of $(\mathcal{X}_n(\omega_\ell))_{\ell=1,\ldots,q}$ when the number of frequencies q grows with the sample size. Derivation of the asymptotic distribution of M_n is already quite delicate when $(X_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence in H. In the scalar case, one can prove that M_n^X is in the domain of attraction of the Gumbel distribution, see Davis and Mikosch (1999).
The purpose of Chapter 3 is to tackle this problem for functional data. Given the complexity of the problem, we restrict to i.i.d. innovations (ε_t). Furthermore, we consider the approximated statistic $M_n^p = \max_{j=1,\ldots,q} \|\mathcal{X}_n^p(\omega_j)\|^2$, where $\mathcal{X}_n^p(\omega_j)$ is the p–dimensional approximation of $\mathcal{X}_n^p(\omega)$ via the Karhunen–Loève expansion. We have showed successively that the three following statistics: M_n^p, $M_n^{p_n}$ where $p = p_n \nearrow +\infty$ and M_n are in the attration domain of the Gumbel distribution. The first result generalises the result of Davis and Mikosch (1999) in the multivariate setting, the second specify at which rate we can let the dimension $p = p_n$ grow, and the third provides, under some further assumptions on the distribution of X_t, a trully functional generalisation of Davis and Mikosch (1999). Let $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ be some normalising sequences (to be defined later). In the first part of the proof we will use that

$$a_n^{-1}(M_n - b_n) = a_n^{-1}(M_n - M_n^p) \; + \; a_n^{-1}(M_n^p - b_n)\,, \tag{1.19}$$

and show that the first term converges to zero in probability provided that p grows "sufficiently" fast. The growth rate of p depends on the decay rate of the eigenvalues of $C = \mathrm{Var}(X_1)$. For the second term, we proceed as follows: it trivially holds that

$$\left| P\left(a_n^{-1}(M_n^p - b_n) \leq x\right) - e^{-e^{-x}} \right| \leq \rho_{n,p} + \left| P\left(a_n^{-1}(\widetilde{M}_n^p - b_n) \leq x\right) - e^{-e^{-x}} \right|,$$

where $\widetilde{M}_n^p$ is defined as M_n^p, but based on i.i.d. Gaussians elements $Y_1, \ldots, Y_n$ with same covariance operator C as X_t, and $\rho_{n,p} = \sup_{x\in\mathbb{R}} |P(M_n^p \leq x) - P(\widetilde{M}_n^p \leq x)|$. To prove that the second term converges to zero, we first remark that for a fixed $p \geq 1$, $\|\mathcal{X}_n^p(\omega)\|^2$ is distributed as an hypoexponential random variable, which is already known to belong to the Gumbel domain of attraction, see e.g. Kang and Serfolo (1999). We show in Lemma 13 that this still holds when $p = o(n)$ and for $a_n = \lambda_1$ and $b_n = \lambda_1 \log(n/2) - \lambda_1 \sum_{j=2}^{\infty} \log(1 - \lambda_j/\lambda_1)$, where $(\lambda_j)_{j\geq}$ are the decreasingly ordered eigenvalues of C.

In order to get a bound for $\rho_{n,p}$, we use a brilliant result from Chernozhukov et al. (2017). They provide a Gaussian approximation of $P(n^{-1/2}\sum_{i=1}^n \xi_i \in A)$, where $\xi_1, \ldots, \xi_n$ are i.i.d. random vectors, and A runs through the class of s–sparsely convex subsets of $\mathbb{R}^d$. These are intersections of sets whose indicator function depends on at most s–components. This result mainly relies on a sharp anti–concentration inequality due to Nazarov (2003). In our situation, we deduce that

$$\rho_{n,p} \;\leq\; \kappa \cdot \frac{p^3 \log(n)}{\lambda_p^{1/2} n^{1/6}}, \quad \text{for some constant } \; \kappa > 0.$$

Note that the presence of λ_p in the denominator is only a problem in the functional setting, i.e. when we let p go to infinity and thus, λ_p go to zero. Using this bound, we can show that under some restrictions on the distribution of X_t, there exists a sequence of integers $p = p_n$, such that both terms in (1.19) converge, namely that the first one converges to zero in probability whereas the second one converges in distribution to the Gumbel distribution.

Note that our results generalise the result of Davis and Mikosch (1999) for univariate time series, to the multivariate and functional setting. In finite dimension, we also investigate the non i.i.d. case, i.e. when $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a linear process. Finally, we describe how to extend the test in Hörmann et al. (2017) using our results.

Functional GARCH

In Chapter 4, we propose a generalisation of the celebrated GARCH model to the functional setting. Our model, shown already above in (1.8), is similar to the one of Hörmann et al. (2013) for the ARCH(1) and Aue et al. (2016) for the GARCH(1,1). Moreover, we also introduce higher ordered models, i.e. functional GARCH(p,q) for $p \geq 1$ or $q \geq 1$. Staying for the moment in the GARCH(1,1) framework, we assume that the operators $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are kernel operators, and that

$$\sigma_t^2(u) = \delta(u) + \int_0^1 K_{\boldsymbol{\alpha}}(u,v) y_{t-1}^2(v)dv + \int_0^1 K_{\boldsymbol{\beta}}(u,v)\sigma_{t-1}^2(v)dv, \quad \forall u \in [0,1], \tag{1.20}$$

where $\delta(u) > 0$, and $K_{\boldsymbol{\alpha}}(u,v)$, $K_{\boldsymbol{\beta}}(u,v) \geq 0$, for all u, $v \in [0,1]$. Recall that $y_t(u) = \sigma_t(u)\eta_t(u)$, where the innovations process $(\eta_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence in $L^2[0,1]$, thus, the volatility curves can be interpreted as the pointwise volatility since we have that

$$P\Big(|y_t(u)| < \sigma_t(u) \cdot Q_{1-\alpha/2}^{\eta(u)} \,\Big|\, y_s,\ s < t\Big) = 1 - \alpha, \quad \forall u \in [0,1].$$

Here Q_α^X is the α-quantile of the variable X.
Similarly as in the scalar GARCH, the coefficient (operator) $\boldsymbol{\alpha}$, induces the sensibility of the volatility to shocks since it produces the dependence between the current volatility curve σ_t^2 and the curve y_{t-1}^2. The coefficient (operator) $\boldsymbol{\beta}$ induces persistence of the volatility, since it represents the autoregressive dependence of the volatility process. We illustrate these phenomenons on Figure 8, which displays a simulation of our model.
We use a Gaussian process for the innovations, and thus from the identifiability assumption, i.e. $E[\eta^2(u)] = 1$ for all $u \in [0,1]$, we get that $Q_{1-\alpha/2}^{\eta(u)} \approx 2$, for the level $\alpha = 0.05$.
We then show that a sufficient condition for the existence of a non-anticipative ergodic and strictly stationary solution $(y_t)_{t\in\mathbb{Z}}$ to (1.8), is that $\gamma < 0$, where

$$\gamma = \lim_{t\to\infty} \frac{1}{t} \log \|\Psi_t \dots \Psi_0\| \quad \text{ and } \quad \Psi_t : \begin{cases} L^2[0,1] & \to \ L^2[0,1] \\ x & \mapsto \ \boldsymbol{\alpha}(x \cdot \eta_t^2) + \boldsymbol{\beta}(x). \end{cases}$$

Our condition is milder than the one recently obtained in Aue et al. (2016), namely $E \log \|\Psi_0\|_{\mathcal{S}} < 0$. Note that in finite dimension, the definition of the so–called *Lyapounov top exponent* γ associated to a sequence of random matrices $(\Psi_t)_{t\in\mathbb{Z}}$, is independent of the matrix

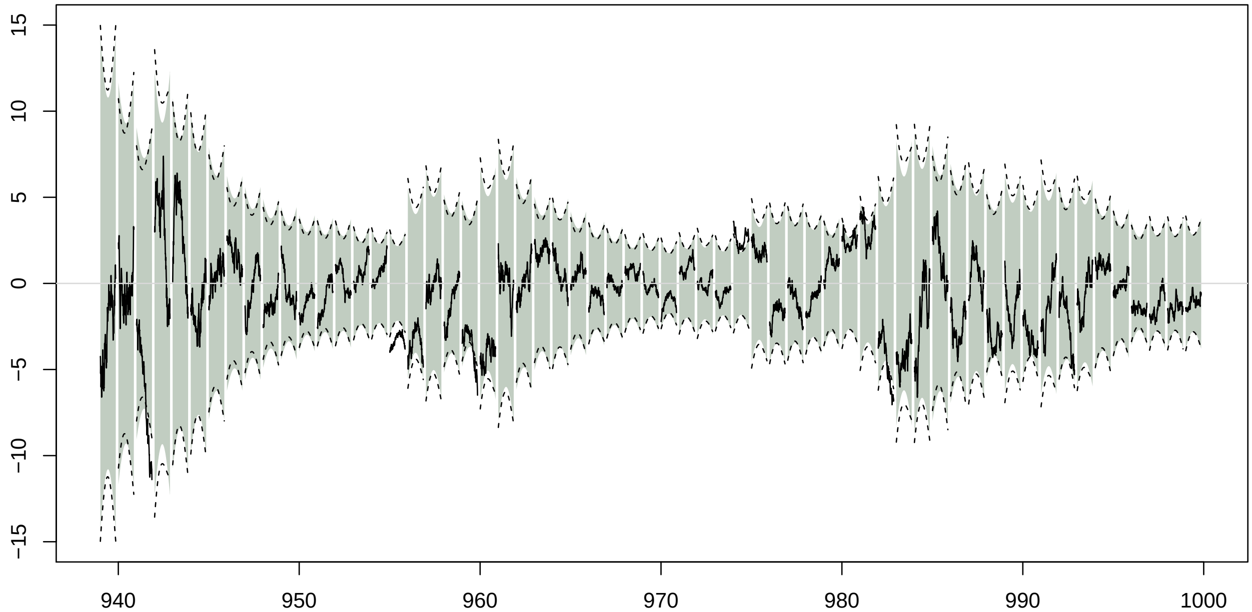

Figure 8: Solid lines represent the simulated process y_t, the shaded area is the region $\{[-2\sigma_t(u), 2\sigma_t(u)] \colon u \in [0,1]\}$. The dashed lines are estimators $\pm 2\widetilde{\sigma}_t(\hat{\theta})(u)$.

norm used. This is not the case in our situation, since norms are not necessarily equivalent in $L^2[0,1]$.

For the estimation, we have first attempted a method of moments approach, similar as the one in Hörmann et al. (2013) for the functional ARCH. Note that from (1.8) we can deduce that the squares of a functional GARCH satisfy

$$y_t^2 - (\boldsymbol{\alpha} + \boldsymbol{\beta})(y_{t-1}^2) = \delta + \eta_t - \boldsymbol{\beta}(\eta_{t-1}), \tag{1.21}$$

where $\eta_t = y_t^2 - \sigma_t^2$ is a *martingale difference sequence* (MDS). In other words, we can transform the estimation problem to the *functional ARMA* context, with MDS innovations. However, estimation of functional ARMA models is still not tackled adequately in the literature and so we first tried to develop a fully functional method to estimate the functional moving average coefficient Θ in (1.6). Unfortunately this approach did not lead to satisfactory results and seemed overly difficult, see in the Appendix 5.2. Note that recent results of Aue and Klepsch (2017b) in functional MA estimation look promising and could lead to potential breakthrough for this approach.

Still, in Chapter 4 we have decided to take an alternative route. We decided to parametrise our model, i.e. we identify $(\delta, \boldsymbol{\alpha}, \boldsymbol{\beta})$ with an element θ in some compact space Θ. This is also the approach in Aue et al. (2016). However, they used a *least squares estimator* whose asymptotic normality necessitates eighth order moments and moreover, from classical GARCH literature it is known that it cannot compete likelihood methods, which are commonly used in this context. This is why we decided to pursue a *quasi maximum likelihood* approach. Clearly, in the infinite dimensional setting we cannot (at least not in a straight forward manner) define a likelihood function. To this end, we consider some instrumental non-negative functions

$\varphi_1, \ldots, \varphi_M$ in $L^2[0,1]$, and define our estimator $\widehat{\theta}_n$ as the minimiser over Θ of the following criterion

$$\widetilde{Q}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \tilde{\ell}_t(\theta), \quad \tilde{\ell}_t(\theta) = \sum_{m=1}^{M} \left\{ \frac{\langle y_t^2, \varphi_m \rangle}{\langle \tilde{\sigma}_t^2, \varphi_m \rangle} + \log\langle \tilde{\sigma}_t^2, \varphi_m \rangle \right\},$$

where the *empirical volatility* $\tilde{\sigma}_t^2$ is computed recursively through a similar equation such as (1.20) and some initial values y_0 and $\tilde{\sigma}_0$ in $L^2[0,1]$. Under the strong assumption that the parameters δ, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ have finite rank, we show that our process can be related to some multivariate semi–strong CCC-GARCH, i.e. a CCC-GARCH whose innovation process is not i.i.d. but is rather an MDS. Using this we provide in Theorems 13 and 14 the consistency and asymptotic normality of $\widehat{\theta}_n$. As a side result we prove in Theorem 15 and Theorem 16 the consistency and asymptotic normality of the QML estimator for semi–strong CCC-GARCH processes.

In a next step we have weakened our assumptions and showed in Proposition 7 that our estimation method can be adapted to a truly functional setting, i.e. without assuming that the parameters δ, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are of finite rank. Here we were not able to use some underlying multivariate GARCH process, but we managed to prove the consistency directly by using an infinite set of functions φ_m's to identify the true parameter. Then, since the parameter space Θ is compact, we were able to apply a technical result from Section 2 i.e. Lemma 6. Finally we supported the relevance of our approach by simulations and illustrated it on some real data.

2 On the CLT for discrete Fourier transforms of functional time series

Clément Cerovecki* and Siegfried Hörmann†

Abstract

Abstract. The purpose of this paper is to derive sharp conditions under which the discrete Fourier transform

$$\mathcal{X}_n(\omega) := n^{-1/2} \sum_{t=1}^{n} X_t e^{-\mathrm{i}t\omega}, \quad \omega \in (-\pi, \pi],$$

of a functional time series $(X_t)_{t\in\mathbb{Z}}$ is asymptotically normal. Assuming that the function space is a Hilbert space we prove that a central limit theorem (CLT) holds for almost all frequencies ω if the process $(X_t)_{t\in\mathbb{Z}}$ is stationary, ergodic and purely non-deterministic. Under slightly stronger assumptions we formulate versions which provide a CLT for fixed frequencies as well as for $\mathcal{X}_n(\omega_n)$, when $\omega_n \to \omega_0$ is a sequence of fundamental frequencies. In particular we also deduce the regular CLT ($\omega = 0$) under new and very mild assumptions. We show that our results apply to the most commonly studied functional time series.

Keywords: central limit theorem, functional time series, Fourier transform, periodogram, stationarity

2.1 Introduction

Functional time series analysis is a branch of the emerging statistical field of functional data analysis (FDA)—we refer to the monographs Ramsay and Silverman (2006); Ferraty and Vieu (2006); Horváth and Kokoszka (2012); Hsing and Eubank (2012). A quickly accessable recent overview is presented in Cuevas (2014). An excellent literature survey for recent developments in FDA can be found in Goia and Vieu (2016). This paper serves as the introduction to a special volume of JMVA which contains a number of new contributions to the field.
The need for functional time series (FTS) methodology is easily explained by the fact that many functional data are sequentially sampled and serially correlated by their very nature. A common situation is that a continuous time process is cut into natural segments, such as days. Then there is not just dependence within the individual curves but also across curves and we obtain a time series $(X_t)_{t\in\mathbb{Z}}$ with realizations in some function space, i.e., every

*Université libre de Bruxelles, Boulevard du Triomphe, CP210 B-1050 Brussels - Belgium Office: NO O.9.211. *E-Mail address*: clement.cerovecki@ulb.ac.be

†Université libre de Bruxelles, Boulevard du Triomphe, CP210 B-1050 Brussels - Belgium Office: NO O.9.115 *shormann@ulb.ac.be*

observation X_t is a random curve $(X_t(\tau)\colon \tau \in \mathcal{U})$ with some continuous domain $\mathcal{U}$. Despite the fact that there exist a variety of recent results related to forecasting functional time series data (e.g., Hyndman and Shang (2009); Aue et al. (2015); Ruiz-Medina et al. (2016); Liebl (2013); Guillas et al. (2011); Kowal et al. (2016); Klepsch and Klüppelberg (2017)), most of the available FDA methodology is devoted to i.i.d. samples. Functional time series methodology in general has not yet received as much attention as one might expect. One of the first and most seminal contributions is Bosq (2000). This monograph is formulating the basic theoretical foundation for FTS. The book focuses on the analysis of functional AR (FAR) processes, which are nowadays among the most popular and best studied FTS models. A core reason for this is that these processes are very convenient for prediction (see e.g. Kowal et al. (2016); Aue et al. (2015); Guillas et al. (2011); Kargin and Onatski (2008); Didericksen et al. (2012)). Further refinements related to FAR modeling can e.g. be found in Damon and Guillas (2005); Mas (2007).

The FAR setting also provides a structural dependence framework. For asymptotic inference on time series data imposing a certain dependence assumption is crucial. Typically some form of near-epoch dependence or mixing assumptions, like strong mixing or cumulant mixing, are used for FTS in order establish large sample results (see e.g. Ferraty et al. (2002); Hörmann and Kokoszka (2010); Panaretos and Tavakoli (2013a).) In this paper we will work with so-called *purely non-deterministic processes* (see Assumption 1 below). This framework is very general and includes many commonly employed weak-dependence frameworks.

In our paper, like in most other time series contributions, stationarity is a crucial assumption. To verify this assumption for functional data we refer to Horváth et al. (2014), who propose corresponding tests.

Lately, there have been some papers devoted to frequency domain analysis for FTS, e.g., Panaretos and Tavakoli (2013a); Hörmann et al. (2015a,b). In contrast to the time domain analysis, which is based on analyzing the data sample and the auto-covariance function, the frequency domain analysis is grounded on the *discrete Fourier transform (DFT)*

$$\mathcal{X}_n(\omega) = n^{-1/2} \sum_{t=1}^{n} X_t e^{-\mathrm{i}t\omega}, \quad \omega \in (-\pi, \pi]$$

of some FTS $(X_t)_{t\in\mathbb{Z}}$ and its *spectral density operator*,

$$\mathcal{F}_\omega := \sum_{h\in\mathbb{Z}} C_h e^{-\mathrm{i}h\omega}. \tag{2.1}$$

Here C_h is the lag h covariance operator of the stationary functional time series—the precise definition of C_h is given below in Section 2.2. The sample $X_1, \dots, X_n$ and $(C_h)_{h\in\mathbb{Z}}$ are equivalent to $(\mathcal{X}_n(2\pi\ell/n))_{\ell=-\lfloor\frac{n-1}{2}\rfloor,\dots,\lfloor\frac{n}{2}\rfloor}$ and $(\mathcal{F}_\omega\colon \omega \in (-\pi,\pi])$, respectively, in the sense that one can be obtained from the other. Depending on the nature of the problem, one or the other approach may be simpler or more effective. For example, Panaretos and Tavakoli (2013b) and Hörmann et al. (2015a) demonstrate that dimension reduction via principal components is more effective (in fact optimal in a certain sense) when done in the frequency domain.

Like for scalar or multivariate time series, the DFT is the main building block for the frequency domain analysis, and hence understanding its asymptotics is a fundamental problem. Moreover, the DFT is of direct interest to statisticians since it is closely related to the *periodogram* which can, for example, be used to detect some underlying periodic behavior of the time series. (See, e.g., Brockwell and Davis (1991).) In Hörmann et al. (2017) a variety of test statistics which can be used to reveal a periodic trend in a FTS are proposed and examined. Those tests involve in one or the other way the functional DFT as a building block. It is demonstrated, for example, that a standard functional ANOVA test—which can be used for this problem—is composed in such a way. Unless $(X_t)_{t\in\mathbb{Z}}$ is a Gaussian process, the exact distribution of the DFT is unreachable and hence, in order to derive critical values for the tests, a Fourier CLT is needed. Let us also mention that very recently a bootstrap procedure to approximate the distribution of $\mathcal{X}_n(\omega)$ has been developed (Paparoditis (2016)). Clearly, when $\omega = 0$ we obtain the regular partial sums process, which is without any doubt of fundamental importance for statistical inference. In Horváth et al. (2013) asymptotic normality under L^2–m-approximability (see Section 2.3.3) is derived and used to test for equality in mean of two time series samples. This CLT is a special case of our Theorem 7 below.

In a seminal paper Panaretos and Tavakoli (2013a) have shown that under regularity assumptions $\mathcal{X}_n(\omega)$ converges to a (complex) Gaussian random element with covariance operator $\mathcal{F}_\omega$. (The actual definition of $\mathcal{F}_\omega$ involves scaling by $\frac{1}{2\pi}$, which we omit here.) Hence, $\mathcal{F}_\omega$ can be interpreted as the asymptotic covariance operator of the discrete Fourier transform (DFT). In Panaretos and Tavakoli (2013a) it is assumed that $\sum_{h\in\mathbb{Z}} \|C_h\|_\mathcal{T} < \infty$ (here $\|\cdot\|_\mathcal{T}$ denotes the trace norm—see Section 2.2) in order to assure convergence of the series in (2.1). It follows that $\mathcal{F}_\omega$ is a nuclear operator, i.e., it has a finite trace. This is an important feature when it comes to verifying tightness of $(\mathcal{X}_n(\omega))_{n\geq 1}$. Regarding the dependence structure, a cumulant type mixing condition for functional data is used. The nice feature of such mixing conditions is that no specific time series model needs to be imposed. Still, this approach requires to compute and bound functional cumulants of all orders, which is generally not an easy task and necessitates moments of all orders. *The main objective of this paper is then to relax these conditions. All our theorems below hold assuming only finite second moments and some very general form of weak dependence.*

For real valued processes asymptotic normality for $\mathcal{X}_n(\omega)$ has been obtained under several dependence conditions. Here we only cite the early paper of Walker (1965) who considered linear processes, a survey article of Kokoskza and Mikosch (2000) and the more recent contributions of Wu (2004) and Peligrad and Wu (2010). The latter paper covers a variety of special cases, including strong mixing sequences. It also contains a more detailed literature survey. One of the main results of our article is an extension of the CLT of Peligrad and Wu (2010) to functional data. We show the weak convergence of $\mathcal{X}_n(\omega)$ for *purely non-deterministic* processes. More precisely, letting $\mathcal{G}_t = \sigma(X_t, X_{t-1}, \ldots)$—the σ-algebra generated by $(X_s)_{s\leq t}$—and $\mathcal{G}_{-\infty} = \bigcap_{t\geq 0} \mathcal{G}_{-t}$ we impose the following assumption.

Assumption 1. *The process $(X_t)_{t\in\mathbb{Z}}$ is stationary and ergodic and satisfies $E[X_0|\mathcal{G}_{-\infty}] = 0$ a.s.*

We remark that a conditional expectation for random elements in Hilbert spaces as just stated is well defined if $E\|X_0\| < \infty$ (see, e.g., (Bosq, 2000, p.29)). Besides the obligatory existence of second order moments, Assumption 1 will be the only condition needed for the CLT presented below in Theorem 1. Since we will not impose any further condition ensuring summability of the C_h, a tricky part is the construction and definition of the spectral density

operator. Our construction will be an indirect one based on a completeness argument in an appropriate Hilbert space.
In Theorem 2 we will give a result which is slightly less general, but is more useful in applications since it will allow for more explicit constructions of $\mathcal{F}_\omega$. In our Theorem 3 we consider the case $\omega = 0$ and derive the CLT for regular partial sums. These main results along with the precise technical setting are presented in Section 2.2. In Section 2.3 we consider application of our theorems to so-called Bernoulli shifts. Within this framework we can further refine the asymptotics and consider the weak convergence of $S_n(\omega_n)$ when $(\omega_n)_{n\geq 1}$ is a convergent sequence of *fundamental frequencies*. We also show how the theorems apply in some commonly employed dependence frameworks for functional time series models and compare the required conditions to existing ones in the literature. Proofs are given in Section 2.4.

2.2 Main results

We start by introducing further notation and stating the setup precisely.
The process $(X_t)_{t\in\mathbb{Z}}$ is defined on some probability space $(\Omega, \mathcal{A}, P)$ and takes values in some real *separable Hilbert space* H_0. Although our observations are assumed to be real, the very definition of $\mathcal{X}_n(\omega)$ necessitates to adopt a complex setting. So we will henceforth consider the complex Hilbert space $H = H_0 + \mathrm{i}H_0$. Throughout u denotes a generic element in H. Then u is of the form $u = u_0 + \mathrm{i}u_1$ where u_0, u_1 denote generic elements in H_0. We denote $\overline{u} = u_0 - \mathrm{i}u_1$. The space H is equipped with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$ which are induced from the inner product $\langle \cdot, \cdot \rangle_{H_0}$ on H_0, i.e., $\langle u_0 + \mathrm{i}u_1, v_0 + \mathrm{i}v_1 \rangle = \langle u_0, v_0 \rangle_{H_0} + \langle u_1, v_1 \rangle_{H_0} + \mathrm{i}\Big(\langle u_1, v_0 \rangle_{H_0} - \langle u_0, v_1 \rangle_{H_0}\Big)$. We write $X \in L^p_H(\Omega)$ (short for $X \in L^p_H(\Omega, \mathcal{A}, P)$) to indicate that $E\|X\|^p < \infty$. The space $L^p_H(\Omega)$ is a Banach space and for $p = 2$ again a Hilbert space with inner product $E\langle X, Y \rangle$. The covariance operator $\mathrm{Cov}(X, Y)$ on $L^2_H(\Omega)$ is defined as $\mathrm{Cov}(X, Y)(u) = E\Big[(X - EX)\langle u, Y - EY \rangle\Big]$ and $\mathrm{Var}(X) := \mathrm{Cov}(X, X)$. We denote by $C_h = \mathrm{Cov}(X_h, X_0)$, $h \in \mathbb{Z}$, the lag h autocovariance operator of the time series. Expectations or other integrals for elements with values in Banach spaces are understood in the sense of Bochner integrals, see, e.g., Mikusiński (1978).
In the following $\mathcal{N}_{H_0}(\mu, \Sigma)$ denotes a Gaussian element in H_0 with mean μ and covariance operator Σ. Then $X \sim \mathcal{N}_{H_0}(\mu, \Sigma)$ if and only if the projection $\langle X, u_0 \rangle$ is normally distributed with mean $\langle \mu, u_0 \rangle$ and variance $\langle \Sigma(u_0), u_0 \rangle$. A complex $Z = Z_0 + \mathrm{i}Z_1 \in H$ is said to be Gaussian if (Z_0, Z_1) is a Gaussian element in $H_0 \times H_0$. Define $\mu_i = EZ_i$ $(i = 0, 1)$, $V_{ij} = \mathrm{Cov}(Z_i, Z_j)$ $(i, j \in \{0, 1\})$ and set $\mu = \mu_0 + \mathrm{i}\mu_1$. Moreover, set $\Gamma = \mathrm{Var}(Z)$ and let $C(u) = E\Big[(Z - \mu)\langle u, \overline{Z - \mu} \rangle\Big]$ be the *relation operator* of Z. By simple algebra $\Gamma(u) = V_{00}(u) + V_{11}(u) + \mathrm{i}\Big(V_{10}(u) - V_{01}(u)\Big)$ and $C(u) = V_{00}(u) - V_{11}(u) + \mathrm{i}\Big(V_{01}(u) + V_{10}(u)\Big)$. With $\mathrm{Re}(\Gamma) := V_{00} + V_{11}$ and $\mathrm{Im}(\Gamma) := V_{10} - V_{01}$ and analogue definitions for $\mathrm{Re}(C)$ and $\mathrm{Im}(C)$ it follows then that

$$\begin{pmatrix} Z_0 \\ Z_1 \end{pmatrix} \sim \mathcal{N}_{H_0 \times H_0}\left(\begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}, \frac{1}{2}\begin{bmatrix} \mathrm{Re}(\Gamma + C) & -\mathrm{Im}(\Gamma - C) \\ \mathrm{Im}(\Gamma + C) & \mathrm{Re}(\Gamma - C) \end{bmatrix}\right). \tag{2.2}$$

Relation (2.2) implies that the law of Z is determined by μ, Γ and C. We write $Z \sim \mathcal{CN}_H(\mu, \Gamma, C)$. It can be readily shown that $Z \sim \mathcal{CN}_H(0, \Gamma, C)$, if and only if for any $u \in H$

we have

$$\langle Z, u\rangle \sim \mathcal{CN}_{\mathbb{C}}\Big(0, \langle \Gamma(u), u\rangle, \langle C(\overline{u}), u\rangle\Big).$$

In this paper we are mainly dealing with the *circularly-symmetric* case of Gaussian elements, i.e., when $\mu = 0$ and $C = 0$ (see, e.g., (Brockwell and Davis, 1991, p. 444) for the multivariate case). The other important special case is if Z is real (i.e., when $Z_1 = 0$). This is equivalent to say that Γ is real and equals C. Then we can either consider Z as element in H_0 with $Z \sim \mathcal{N}_{H_0}(0, \Gamma)$ or view it as an element in H with $Z \sim \mathcal{CN}_H(0, \Gamma, \Gamma)$. We take the latter point of view.

Below we will consider bounded linear, compact operators $A : H \to H$. Recall that for some orthonormal basis (ONB) $(v_j)_{j\geq 1}$ of H the Hilbert-Schmidt norm of A is $\|A\|_{\mathcal{S}} = \Big(\sum_{j\geq 1}\|A(v_j)\|^2\Big)^{1/2}$ and the Trace norm of A is $\|A\|_{\mathcal{T}} = \sum_{j\geq 1}\langle (A^*A)^{1/2}(v_j), v_j\rangle$. Both norms are independent of the choice of $(v_j \colon j \geq 1)$. If $\|A\|_{\mathcal{S}} < \infty$ we say that A is Hilbert-Schmidt and if $\|A\|_{\mathcal{T}} < \infty$ we say that A is trace class. We have $\|A\|_{\mathcal{S}} \leq \|A\|_{\mathcal{T}}$. If A is self-adjoint and non-negative definite then $\operatorname{tr}(A) = \|A\|_{\mathcal{T}} = \sum_{j\geq 1}\langle A(v_j), v_j\rangle$. For a zero mean element $X \in L^2_H(\Omega)$ it holds that $\operatorname{tr}\Big(\operatorname{Var}(X)\Big) = E\|X\|^2$. Finally we recall that a sequence of operators A_n on H is said to converge in the *weak operator topology* to A if $\langle A_n(u), v\rangle \to \langle A(u), v\rangle$ for all $u, v \in H$. Short we write $A_n \xrightarrow{\text{w}} A$.

Theorem 1. *Let $(X_t)_{t\in\mathbb{Z}}$ be a sequence in $L^2_{H_0}(\Omega)$ which satisfies Assumption 1. Then for almost every $\omega \in (-\pi, \pi]$ there exists a linear operator $\mathcal{F}_\omega$, which is self-adjoint and non-negative definite such that*

$$\mathcal{X}_n(\omega) \xrightarrow{d} \mathcal{CN}_H\left(0, \mathcal{F}_\omega, 0\right).$$

Moreover we have that

(I) $\operatorname{Var}\Big(\mathcal{X}_n(\omega)\Big) \xrightarrow{\text{w}} \mathcal{F}_\omega$;

(II) $E\|\mathcal{X}_n(\omega)\|^2 = \operatorname{tr}\Big(\operatorname{Var}\Big(\mathcal{X}_n(\omega)\Big)\Big) \to \operatorname{tr}\Big(\mathcal{F}_\omega\Big) < \infty$;

(III) $C_h = \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathcal{F}_\omega\, e^{\mathrm{i}h\omega}d\omega, \qquad \forall h \in \mathbb{Z}$;

(IV) *for almost all $(\omega, \omega') \in (-\pi, \pi]^2$ the components of $(\mathcal{X}_n(\omega), \mathcal{X}_n(\omega'))$ are asymptotically jointly Gaussian and independent.*

We call $\mathcal{F}_\omega$ the spectral density operator of $(X_t)_{t\in\mathbb{Z}}$ and remark that it is generally not explicitly defined as in (2.1). In fact, the series in (2.1) may not be convergent under our mild assumptions. Since

$$\mathcal{F}_{n;\omega} := \operatorname{Var}\Big(\mathcal{X}_n(\omega)\Big) = \sum_{|h|<n}\left(1 - \frac{|h|}{n}\right)C_h e^{-\mathrm{i}h\omega},$$

relation **(I)** implies solely that the Cesàro averages of $(C_h e^{-\mathrm{i}h\omega})_{h\in\mathbb{Z}}$ converge (in weak operator topology).

For practical reasons it is useful to know for which frequencies Theorem 1 holds. For example, $\omega = 0$ is an important special case, but the theorem doesn't say if this frequency is part of the exceptional null set or not. We will see that one of the delicate steps in the proof of Theorem 1 is to guarantee existence of the operator $\mathcal{F}_\omega$ and to establish the related convergence in **(I)** and **(II)**. By the extremely mild assumptions we are imposing, we can only assure this for almost every ω. Requiring Assumption 2 below allows us to establish the same result for some fixed frequency ω_0. To formulate this assumption we first introduce the projection operator $\mathcal{P}_k := E[\,\cdot\,|\mathcal{G}_k] - E[\,\cdot\,|\mathcal{G}_{k-1}]$, $k \in \mathbb{Z}$. It is elementary that $\mathcal{P}_k$ are linear operators on $L^1_H(\Omega)$. We note at at this point that a key property which we will need is that $\mathrm{Cov}(\mathcal{P}_k(X), \mathcal{P}_\ell(Y)) = 0$ (the zero operator) for all $X, Y \in L^2_H(\Omega)$ when $k \neq \ell$. That is, projections are strongly orthogonal. (See Lemma 2.)

Assumption 2. *The process $(X_t)_{t\in\mathbb{Z}}$ is stationary and ergodic and for some $\omega_0 \in (-\pi, \pi]$ the following properties hold:*

(A1) $Z_n(\omega_0) := \sum_{t=0}^n \mathcal{P}_0(X_t) e^{-\mathrm{i}t\omega_0}$ *is a Cauchy sequence in $L^2_H(\Omega)$;*

(A2) $E\Big\|E[\mathcal{X}_n(\omega_0)|\mathcal{G}_0]\Big\|^2 = o(1)$.

To get some intuition behind **(A1)** and **(A2)** and the general approach we introduce $Z_n^{(k)}(\omega) := \sum_{t=0}^n \mathcal{P}_k(X_{t+k}) e^{-\mathrm{i}t\omega}$ and write

$$\begin{aligned} n^{1/2}\mathcal{X}_n(\omega) &= n^{1/2} \sum_{k=1}^n \mathcal{P}_k(\mathcal{X}_n(\omega)) + n^{1/2} E[\mathcal{X}_n(\omega)|\mathcal{G}_0] \\ &= \sum_{k=1}^n \sum_{t=k}^n \mathcal{P}_k(X_t) e^{-\mathrm{i}t\omega} + n^{1/2} E[\mathcal{X}_n(\omega)|\mathcal{G}_0] \\ &= \sum_{k=1}^n Z_{n-k}^{(k)}(\omega) e^{-\mathrm{i}k\omega} + n^{1/2} E[\mathcal{X}_n(\omega)|\mathcal{G}_0]. \end{aligned} \tag{2.3}$$

The variables $Z_n^{(k)}(\omega)$, $k \geq 1$, are strongly orthogonal and by Assumption **(A1)** they have a limit $Z^{(k)}(\omega)$ in $L^2_H(\Omega)$ for $\omega = \omega_0$. Moreover, it is easy to see that $(Z^{(k)}(\omega))_{k\geq 1}$ is a stationary martingale difference sequence. Together with **(A2)** this guarantees that $n^{1/2}\mathcal{X}_n(\omega)$ is close to $T_n(\omega) := \sum_{k=1}^n Z^{(k)}(\omega) e^{-\mathrm{i}k\omega}$. The partial sum $T_n(\omega)$ is more handy when it comes to study the CLT and to compute the covariance operator. We define

$$\mathcal{F}_\omega := \mathrm{Var}\Big(T_n(\omega)/\sqrt{n}\Big) = \mathrm{Var}\Big(Z^{(1)}(\omega)\Big). \tag{2.4}$$

Theorem 2. *Let $(X_t)_{t\in\mathbb{Z}}$ be a sequence in $L^2_{H_0}(\Omega)$ which satisfies Assumption 2 for some $\omega_0 \in (-\pi, \pi]$. Then*

$$\mathcal{X}_n(\omega_0) \xrightarrow{d} \mathcal{CN}_H\Big(0, \mathcal{F}_{\omega_0}, C_{\omega_0}\Big),$$

where $\mathcal{F}_{\omega_0}$ is defined as in (2.4) *and where $C_{\omega_0} = \mathcal{F}_{\omega_0} I\{\omega_0 \in \{0, \pi\}\}$. Furthermore, the conclusions* **(I)** *and* **(II)** *of Theorem 1 hold for frequency ω_0. If Assumption 2 holds in addition for some $\omega_0' \neq \pm\omega_0$, then conclusion* **(IV)** *of Theorem 1 holds with $(\omega, \omega') = (\omega_0, \omega_0')$.*

As a corollary of this theorem, we obtain the CLT for regular partial sums. It can be phrased as follows.

Theorem 3. *Suppose that $(X_t\colon t \in \mathbb{Z}) \in L^2_{H_0}(\Omega)$. If Assumption 2 holds with $\omega_0 = 0$, then $(X_1 + \cdots + X_n)/\sqrt{n} \xrightarrow{d} \mathcal{N}_{H_0}(0, \mathcal{F}_0)$. We have that $\mathcal{F}_0$ is a non-negative definite, self-adjoint and trace-class operator and that*

$$\sum_{|h|<n} \left(1 - \frac{|h|}{n}\right) C_h \xrightarrow{w} \mathcal{F}_0.$$

We conclude this section with a third and again slightly stronger assumption. Here and in the sequel $\nu_p(X) = \Big(E\|X\|^p\Big)^{1/p}$, $p \geq 1$.

Assumption 3. *The process $(X_t)_{t\in\mathbb{Z}}$ satisfies Assumption 1 and*

(A3) $\sum_{t=0}^{\infty} \nu_2(\mathcal{P}_0(X_t)) < \infty$.

With this new assumption we can obtain the following useful implications.

Lemma 1. *Assumption 3 implies that*

(i) Assumption 2 holds for all $\omega \in (-\pi, \pi]$;

(ii) $\sum_{h\in\mathbb{Z}} \|C_h\|_{\mathcal{S}} < \infty$;

(iii) $\mathcal{F}_\omega$ in (2.1) and (2.4) coincide.

The proof of Lemma 1 is given in Section 2.4.

2.3 Application to Bernoulli shifts

The typical framework we have in mind comprises processes $(X_t)_{t\in\mathbb{Z}}$ which can be represented as Bernoulli shifts, i.e.,

$$X_t = f(\varepsilon_t, \varepsilon_{t-1}, \ldots), \tag{2.5}$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a stationary and ergodic sequence of elements in some normed vector space S and $f : S^{\mathbb{N}} \to H_0$ is measurable. Then $(X_t)_{t\in\mathbb{Z}}$ is stationary and ergodic. We remark that in this case we can use in our theorems the filtration $(\mathcal{G}_k)_{k\in\mathbb{Z}}$ with $\mathcal{G}_k = \sigma(\varepsilon_k, \varepsilon_{k-1}, \ldots)$. Representation (2.5) is very common to many time series models. In particular it applies to the two dependence frameworks we are going to discuss below, namely *linear processes* (possibly with dependent noise) and *$L^2 - m$–approximable processes*. These two concepts cover most of the functional time series models studied in the literature.
When $(\varepsilon_t)_{t\in\mathbb{Z}}$ are i.i.d., then by Kolmogorov's 0-1 law Assumption 1 applies to all such processes. The following convenient condition thus implies Assumption 3.

Assumption 4. *The process $(X_t)_{t\in\mathbb{Z}}$ has representation (2.5) with i.i.d. innovations $(\varepsilon_t)_{t\in\mathbb{Z}}$ and satisfies* **(A3)**.

It should be stressed that $(\varepsilon_t)_{t\in\mathbb{Z}}$ in (2.5) need not necessarily be independent in order to yield Assumption 1. For example, if $(\varepsilon_t)_{t\in\mathbb{Z}}$ are *strongly mixing* then the tail sigma algebra $\mathcal{G}_{-\infty}$ is again trivial (see (Bradley, 2005, p.10)).
For Bernoulli shifts, we can obtain the following refinement of our Theorem 2.

Theorem 4. *Suppose that $(X_t)_{t\in\mathbb{Z}}$ are square integrable random elements satisfying Assumption 4. Suppose that $(\omega_n)_{n\geq 1}$ and $(\omega'_n)_{n\geq 1}$ are two sequences of fundamental frequencies (i.e., $\omega_n \in \frac{2\pi}{n}\mathbb{Z}$) with $\omega_n \to \omega$ and $\omega'_n \to \omega'$. Assume further that for all $n \geq 1$ it holds that $\omega_n \neq \pm\omega'_n$ and $\omega_n, \omega'_n \notin \pi\mathbb{Z}$. Then*

$$(\mathcal{X}_n(\omega_n), \mathcal{X}_n(\omega'_n)) \xrightarrow{d} \mathcal{CN}_{H\times H}\left(0, \begin{bmatrix} \mathcal{F}_\omega & 0 \\ 0 & \mathcal{F}_{\omega'} \end{bmatrix}, 0\right),$$

where $\mathcal{F}_\omega$ is defined as in (2.1).

Note that ω_n and ω'_n are allowed to converge to the same limit. The asymptotic Fourier transforms will stay independent. We also stress that if $\omega_n \to \omega \in \{0, \pi\}$, then $\mathcal{X}_n(\omega_n)$ and $\mathcal{X}_n(\omega)$ have different asymptotics. While for the first we obtain a complex limiting law, the distribution of $\mathcal{X}_n(\omega)$ is real.
In the following subsections we explicitly work out three important types of process and verify for each Assumption 4. We can thus deduce, that Theorems 1, 2, 3 and 4 are applicable for these processes.

2.3.1 Linear processes

Consider a linear process $X_t = \sum_{k\geq 0} \Psi_k(\varepsilon_{t-k})$ where $(\varepsilon_t)_{t\in\mathbb{Z}}$ are i.i.d. and zero mean in some Hilbert space H_1 and $\Psi_k \colon H_1 \to H_0$ are bounded linear operators. We denote by $\|\Psi\|_{\mathcal{L}}$ the operator norm.

Theorem 5. *If $X_t \in L^2_{H_0}(\Omega)$ then Theorem 1 holds. If in addition $\kappa := \sum_{k\geq 0} \|\Psi_k\|_{\mathcal{L}} < \infty$ and $\varepsilon_0 \in L^2_{H_1}(\Omega)$ then Assumption 4 holds. Moreover,*

$$\mathcal{F}_\omega = \Psi(\omega) V \Psi(\omega)^*,$$

where $\Psi(\omega) = \sum_{k\geq 0} \Psi_k e^{-\mathrm{i}k\omega}$ and $\Psi(\omega)^*$ is its adjoint operator and $V = \mathrm{Var}(\varepsilon_0)$.

It is easy to see that $\varepsilon_0 \in L^2_{H_1}(\Omega)$ implies that $X_t \in L^2_{H_0}(\Omega)$. Consequently, our Theorem 5 improves the corresponding result in Panaretos and Tavakoli (2013a), where it is required that $\varepsilon_0 \in L^k_{H_0}(\Omega)$ for all $k \geq 1$.
When $\omega = 0$ we recover the ordinary CLT for the partial sums of $(X_t)_{t\in\mathbb{Z}}$ as, e.g., proven in Merlevède et al. (1997). While for scalar linear processes the CLT only requires square summability of the coefficients, the latter authors prove that in infinite dimensional Hilbert spaces assuming absolute summability is essentially sharp.

Proof. Note that $\mathcal{P}_0(X_t) = \Psi_t(\varepsilon_0)$. Hence, condition **(A3)** follows immediately and thus Assumption 4 holds. The rest follows from the implications of Lemma 1. □

2.3.2 Linear processes with dependent errors

Consider once again a linear process $X_t = \sum_{k\geq 0} \Psi_k(\delta_{t-k})$, where now $(\delta_t)_{t\in\mathbb{Z}}$ has the Bernoulli representation (2.5) $\delta_t = f(\varepsilon_t, \varepsilon_{t-1}, \dots)$ with i.i.d. innovations $(\varepsilon_t)_{t\in\mathbb{Z}}$.

Theorem 6. *Suppose that $(X_t)_{t\in\mathbb{Z}}$ is a linear process as stated above, satisfying the summability condition $\sum_{k\geq 0} \|\Psi_k\|_{\mathcal{L}} < \infty$, and further assume that the process $(\delta_t)_{t\in\mathbb{Z}}$ itself satisfies condition **(A3)**. Then Assumption 4 holds for $(X_t)_{t\in\mathbb{Z}}$.*

For the regular partial sums process, this result compares to Račkauskas and Suquet (2010) who have studied partial sums of linear processes in Banach spaces. They show that the CLT for the innovations transfers to the linear process under summability of $(\|\Psi_k\|_{\mathcal{L}})_{k\geq 0}$.

Proof. We only have to prove condition **(A3)**. It holds that

$$\begin{aligned}\sum_{t\geq 0} \nu_2\Big(\mathcal{P}_0(X_t)\Big) &\leq \sum_{t\geq 0}\sum_{k\geq 0} \nu_2\Big(\mathcal{P}_0(\Psi_k(\delta_{t-k})\Big) \\ &\leq \sum_{k\geq 0} \|\Psi_k\|_{\mathcal{L}} \sum_{t\geq 0} \nu_2\Big(\mathcal{P}_0(\delta_{t-k})\Big) < \infty.\end{aligned}$$

□

2.3.3 $L^2 - m$–approximable processes

Hörmann and Kokoszka (2010) have used the concept of $L^p - m$*–approximability* for analyzing dependent functional data. Then a process $(X_t)_{t\in\mathbb{Z}}$ is said to be $L^p - m$–approximable if X_t has representation (2.5) with i.i.d. innovations and

$$\sum_{m=1}^{\infty} \nu_p(X_0 - X_0^{(m)}) < \infty,$$

where $X_0^{(m)} = f(\varepsilon_0, \dots, \varepsilon_{0-m+1}, \tilde{\varepsilon}_{-m}, \tilde{\varepsilon}_{-m-1}, \dots)$ for some independent copy $(\tilde{\varepsilon}_t\colon t \in \mathbb{Z})$ of $(\varepsilon_t)_{t\in\mathbb{Z}}$. In Hörmann and Kokoszka (2010) it is shown that this concept applies to many stationary and non-stationary functional time series models, including, for example, functional ARCH. The concept is somewhat related to *near epoch dependence (NED)* often employed in the econometrics literature. See, e.g., Pötscher and Prucha (1997). If this condition holds with $p > 2$, then by a recent result of Berkes et al. (2013) a weak invariance principle for the partial sums process holds. This result has been sharpened by Jirak (2013) who proved the same invariance principle under $p = 2$ and also under a milder coupling condition.

Theorem 7. *Suppose that $(X_t)_{t\in\mathbb{Z}}$ is $L^2 - m$–approximable. Then Assumption 4 holds.*

Proof. We first note that under $L^2 - m$-approximability $X_s^{(s)}$ is independent of $\mathcal{G}_0$. First note that the construction yields $\nu^2\Big(\mathcal{P}_0(X_s)\Big) = E\|\mathcal{P}_0(X_s - X_s^{(s)})\|^2$. The right hand side can be bounded by

$$\begin{aligned}&2E\left(\|E[X_s - X_s^{(s)}|\mathcal{G}_0]\|^2 + \|E[X_s - X_s^{(s)}|\mathcal{G}_{-1}]\|^2\right) \\ &\leq 4E\|X_0 - X_0^{(s)}\|^2 = 4\nu_2^2\Big(X_0 - X_0^{(s)}\Big).\end{aligned}$$

□

2.4 Proofs of Chapter 2

Some crucial ideas in our proofs come from Peligrad and Wu (2010). To extend these ideas to the functional setup, several non-trivial steps need to be added. In particular, showing existence of the spectral density operator is delicate. Here it arises as a limiting point in some appropriate Hilbert space—see Lemma 10 and the discussion thereafter. Next, verifying tightness requires extra efforts (Lemmas 4, 6, 7, 8). One step in our proofs is to apply Peligrad and Wu (2010) to the projected Fourier transforms. For the proof of Theorem 1 we need to deal with the fact that each projection will come with its own exceptional set of frequencies where the CLT might fail and one difficulty is to make sure that the exceptional set for the functional DFT is still a null-set (Proposition 1).

2.4.1 Preliminary lemmas

We start with a lemma that discusses basic properties of the projection operators $\mathcal{P}_k$.

Lemma 2. *Let* $X, Y \in L^2_H(\Omega)$*.*

(i) For integers $k \neq \ell$ *we have the strong orthogonality relation*

$$\operatorname{Cov}\Big(\mathcal{P}_k(X), \mathcal{P}_\ell(Y)\Big) = 0;$$

(ii) If X *is* $\mathcal{G}_0$*-measurable, then* $X = \sum_{t \geq 0} \mathcal{P}_{-t}(X) + E[X|\mathcal{G}_{-\infty}]$ *almost surely and in* $L^2_{H_0}(\Omega)$*;*

(iii) If X *is* $\mathcal{G}_0$*-measurable then under Assumption 1*

$$\sum_{t \geq 0} E\|\mathcal{P}_{-t}(X)\|^2 = E\|X\|^2.$$

We notice that *(iii)* can be viewed as Parseval-type identity.

Proof. (i) We have to show that $E\Big[\langle \mathcal{P}_k(X), v\rangle \overline{\langle \mathcal{P}_\ell(Y), u\rangle}\Big] = 0 \quad \forall u, v \in H$. It holds that $\langle \mathcal{P}_k(X), v\rangle = \mathcal{P}_k(\langle X, v\rangle)$ and hence it is enough to restrict to the scalar case. The result follows then straight forwardly from elementary properties for conditional expectations of real valued random variables.

(ii) Note that $\sum_{t=0}^n \mathcal{P}_{-t}(X) = X - E[X|\mathcal{G}_{-n-1}]$. We consider the decaying sequence of σ-algebras $(\mathcal{G}_{-k} \colon k \geq 0)$. For any integrable random variable $X \in H$ the process $(E[X|\mathcal{G}_{-k}] \colon k \geq 0)$ is a reverse martingale with values in H (see, e.g., Chatterji (1964)). It converges a.s. to $E[X|\mathcal{G}_{-\infty}]$. If X is square integrable then convergence also holds in $L^2_{H_0}(\Omega)$.

(iii) By Assumption 1 and *(i)* of this lemma it follows that

$$\sum_{t \geq 0} E\|\mathcal{P}_{-t}(X)\|^2 = \lim_{n \to \infty} E\|X - E[X|\mathcal{G}_{-n}]\|^2 = E\|X\|^2.$$

□

Let $G = L^2_H(\Omega)$ and consider the Hilbert space $L^2_G((-\pi,\pi],\mathcal{B},\lambda)$, with $\mathcal{B}$ and λ being the Borel σ-field and the Lebesgue measure on $(-\pi,\pi]$, respectively. For simplicity we write $L^2_G((-\pi,\pi])$. This space is equipped with inner product $(V,W) = \int_{-\pi}^{\pi} E\langle V(\omega),W(\omega)\rangle d\omega$ and norm $|||V||| = \sqrt{(V,V)}$.

Lemma 3. *Define $Z_n = Z_n(\omega) := \sum_{t=0}^n \mathcal{P}_0(X_t)e^{-\mathrm{i}t\omega}$. Then (Z_n) is a Cauchy sequence in $L^2_G((-\pi,\pi])$, if and only if $\sum_{t\geq 0} E\|\mathcal{P}_0(X_t)\|^2 < \infty$. Moreover, under Assumption 1 the latter summability condition holds.*

We remark that this lemma provides a slightly weaker version of Assumption 2, part 1.

Proof. Using stationarity and the orthogonality of the functions $\omega \mapsto e^{-\mathrm{i}t\omega}$, $\omega \in (-\pi,\pi]$, $(t\in\mathbb{Z})$ we obtain for $m<n$

$$\begin{aligned}|||Z_n - Z_m|||^2 &= \int_{-\pi}^{\pi} E\left\|\sum_{t=m+1}^{n}\mathcal{P}_0(X_t)e^{-\mathrm{i}t\omega}\right\|^2 d\omega = 2\pi\sum_{t=m+1}^{n}E\|\mathcal{P}_0(X_t)\|^2\\ &= 2\pi\sum_{t=m+1}^{n}E\|\mathcal{P}_{-t}(X_0)\|^2.\end{aligned}$$

The result follows by point *(iii)* of Lemma 2. □

It follows under Assumption 1 that there exists an element $Z \in L^2_G((-\pi,\pi])$ with $|||Z_n - Z||| \to 0$. This in turn has some important implications.

(P1) Since

$$|||Z|||^2 = \int_{-\pi}^{\pi} E\|Z(\omega)\|^2 d\omega < \infty,$$

we conclude that $E\|Z(\omega)\|^2 < \infty$ for all $\omega \in M_0 = (-\pi,\pi]\backslash N_0$ where $\lambda(N_0) = 0$. Hence, for all $\omega \in M_0$ the covariance operator $\mathcal{F}_\omega := \mathrm{Var}\big(Z(\omega)\big)$ is well defined, self-adjoint and non-negative definite. The denotation $\mathcal{F}_\omega$ is intentional. As we will see later *it is defining the spectral density operator* (compare to (2.4)). Since $\mathrm{tr}\big(\mathcal{F}_\omega\big) = E\|Z(\omega)\|^2$, this operator is trace class. For $\omega \in N_0$ we set $\mathcal{F}_\omega = 0$.

(P2) There exists a sequence (n_k) such that $E\|Z_{n_k}(\omega) - Z(\omega)\|^2 \to 0$ for all $\omega \in M_1 := (-\pi,\pi]\backslash N_1$, where $\lambda(N_1) = 0$.

(P3) By construction the mapping $\omega \mapsto Z(\omega) \in G$ is measurable, and the mapping from $G \to \mathcal{S}$ (the set of Hilbert-Schmidt operators on H) with $Z(\omega) \mapsto \mathrm{Var}\big(Z(\omega)\big)$, is continuous. Hence, $\omega \to \mathcal{F}_\omega$ is measurable as a mapping from $(-\pi,\pi]$ to the space $\mathcal{S}$, which is known to be a separable Hilbert space. Consequently the integral in **(III)** of Theorem 1 is well defined.

The next lemma will be used in the proof of tightness and implies part **(II)** of Theorem 1.

Lemma 4. *Under Assumption 1 we have for all $\omega \in M_2 = (-\pi,\pi]\backslash N_2$ with $\lambda(N_2) = 0$ that*

$$\mathrm{tr}\big(\mathcal{F}_{n;\omega}\big) = \sum_{|h|<n}\left(1 - \frac{|h|}{n}\right)E\langle X_h, X_0\rangle e^{-\mathrm{i}h\omega} \to \mathrm{tr}\big(\mathcal{F}_\omega\big) < \infty.$$

Proof. Set $c_h := (2\pi)^{-1}\int_{-\pi}^{\pi} E\|Z(\omega)\|^2 e^{\mathrm{i}h\omega}d\omega$. Using (i) we infer from the Fejér-Lebesgue theorem that

$$\sum_{|h|<n}\left(1-\frac{|h|}{n}\right)c_h e^{-\mathrm{i}h\omega} \to E\|Z(\omega)\|^2 = \mathrm{tr}\Big(\mathcal{F}_\omega\Big) < \infty \quad \text{for almost all } \omega.$$

We define M_2 as the set of convergence points. We show now that $c_h = E\langle X_h, X_0\rangle$. Using Lemma 3 and continuity of $|\!|\!|\cdot|\!|\!|$, it can be readily shown that

$$c_h = \lim_{n\to\infty}\frac{1}{2\pi}\int_{-\pi}^{\pi} E\|Z_n(\omega)\|^2 e^{\mathrm{i}h\omega}d\omega.$$

Without loss of generality assume $h \geq 0$. Using stationarity we deduce

$$\begin{aligned}\frac{1}{2\pi}\int_{-\pi}^{\pi} E\|Z_n(\omega)\|^2 e^{\mathrm{i}h\omega}d\omega &= \sum_{t=0}^{n}\sum_{s=0}^{n} E\langle \mathcal{P}_0(X_t), \mathcal{P}_0(X_s)\rangle \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-\mathrm{i}(t-s-h)\omega}d\omega\\ &= \sum_{t=h}^{n} E\langle \mathcal{P}_0(X_t), \mathcal{P}_0(X_{t-h})\rangle = \sum_{t=h}^{n} E\langle \mathcal{P}_{-t}(X_0), \mathcal{P}_{-t}(X_{-h})\rangle.\end{aligned}$$

Since by Lemma 2 *(i)* the terms $\mathcal{P}_{-t}(X_0)$ and $\mathcal{P}_{-s}(X_{-h})$ are orthogonal in $L^2_{H_0}(\Omega)$ for $s \neq t$, it follows that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} E\|Z_n(\omega)\|^2 e^{\mathrm{i}h\omega}d\omega = E\left\langle \sum_{t=h}^{n}\mathcal{P}_{-t}(X_0), \sum_{s=h}^{n}\mathcal{P}_{-s}(X_{-h})\right\rangle.$$

By Assumption 1 and Lemma 2 *(ii)*

$$\sum_{t=h}^{n}\mathcal{P}_{-t}(X_0) = \sum_{t=h}^{n}\mathcal{P}_{-t}(E[X_0|\mathcal{G}_{-h}]) \overset{L^2_{H_0}(\Omega)}{\longrightarrow} E[X_0|\mathcal{G}_{-h}]. \tag{2.6}$$

Similarly, $\sum_{s=h}^{n}\mathcal{P}_{-s}(X_{-h}) \overset{L^2_{H_0}(\Omega)}{\longrightarrow} E[X_{-h}|\mathcal{G}_{-h}] = X_{-h}$. And hence, by continuity of the inner product, $c_h = E\Big\langle E[X_0|\mathcal{G}_{-h}], X_{-h}\Big\rangle = E\langle X_h, X_0\rangle$. □

The next lemma yields property **(III)** of Theorem 1.

Lemma 5. *The operators $\mathcal{F}_\omega$ define the* spectral density operators *of $(X_t\colon t\in\mathbb{Z})$ at frequency ω. This is*

$$C_h = \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathcal{F}_\omega\, e^{\mathrm{i}h\omega}d\omega, \qquad \forall h\in\mathbb{Z}.$$

Proof. We have seen in (P3) that the mapping $\omega \mapsto \mathcal{F}_\omega$ is measurable. The integrand is valued in the separable Hilbert space $\mathcal{S}$. Since $\int_{-\pi}^{\pi}\|\mathcal{F}_\omega\|_{\mathcal{S}}d\omega \leq \int_{-\pi}^{\pi}\mathrm{tr}\Big(\mathcal{F}_\omega\Big)d\omega < \infty$ we know that $\mathcal{F}_\omega$ is strongly integrable and hence we can define (in the sense of a Bochner integral) $I = \int_{-\pi}^{\pi}\mathcal{F}_\omega\, e^{\mathrm{i}h\omega}d\omega$. Let $u, v \in H$. Since Bochner integrals are interchangeable with bounded linear operators we obtain

$$\begin{aligned}\langle I(v), u\rangle &= \int_{-\pi}^{\pi}\langle \mathcal{F}_\omega(v), u\rangle e^{\mathrm{i}h\omega}d\omega = \int_{-\pi}^{\pi} E\Big[\langle Z(\omega), u\rangle\overline{\langle Z(\omega), v\rangle}\Big]e^{\mathrm{i}h\omega}d\omega\\ &= \lim_{n\to\infty}\int_{-\pi}^{\pi} E\Big[\langle Z_n(\omega), u\rangle\,\overline{\langle Z_n(\omega), v\rangle}\Big]e^{\mathrm{i}h\omega}d\omega.\end{aligned}$$

The last equality can be deduced from $\|\|Z_n - Z\|\| \to 0$. Assume now without loss of generality that $h \geq 0$. Similar arguments as in Lemma 4 lead to

$$\frac{1}{2\pi}\langle I(v), u\rangle = \lim_{n\to\infty} E\Big[\Big\langle \sum_{t=0}^{n-h} \mathcal{P}_{-t}(X_h), u\Big\rangle \Big\langle v, \sum_{s=0}^{n-h} \mathcal{P}_{-s}(X_0)\Big\rangle\Big].$$

From (2.6) it follows that

$$\frac{1}{2\pi}\langle I(v), u\rangle = E\Big[\langle E[X_h|\mathcal{G}_0], u\rangle \; \langle v, X_0\rangle\Big] = \langle C_h(v), u\rangle.$$

Since u and v are arbitrary in H, we can infer that $I = 2\pi C_h$. □

We show next that the projections $\langle \mathcal{X}_n(\omega), u\rangle$, $u \in H$, converge weakly to $\langle S_0(\omega), u\rangle$, where $S_0(\omega) \sim \mathcal{CN}_H(0, \mathcal{F}_\omega, 0)$ is the limiting complex Gaussian element and where $\mathcal{F}_\omega$ is defined as in (P1). The first step towards this result is given by the following proposition.

Proposition 1. *Under Assumption 1 there exists for all $u \in H$ a set $\widetilde{N} \subset (-\pi, \pi]$ with $\lambda(\widetilde{N}) = 0$, such that on $\widetilde{M} = (-\pi, \pi]\backslash\widetilde{N}$ the following holds:*

(a) $\lim_{n\to\infty} \mathrm{Var}(\langle \mathcal{X}_n(\omega), u\rangle) = \lim_{n\to\infty}\langle \mathcal{F}_{n;\omega}(u), u\rangle = \langle \mathcal{F}_\omega(u), u\rangle$*;*

(b) $\langle \mathcal{X}_n(\omega), u\rangle \xrightarrow{d} \mathcal{CN}_{\mathbb{C}}(0, \langle \mathcal{F}_\omega(u), u\rangle, 0)$*;*

Proof. We first show that there exists for all $u \in H$ a set $N_u \subset (-\pi, \pi]$ with $\lambda(N_u) = 0$, such that on $M_u = (-\pi, \pi]\backslash N_u$ **(a)** and **(b)** hold.
Let $u \in$ and let $\mathcal{G}_k(u)$ be the filtration of the process $(\langle X_t, u\rangle\colon t \in \mathbb{Z})$. From the results in Peligrad and Wu (2010) we obtain that $\mathrm{Var}(\langle \mathcal{X}_n(\omega), u\rangle) \to f^u(\omega)$ for some function $f^u(\omega)$ which is finite on M_u. More precisely, slightly adapting the proofs of Lemmas 4.1. and 4.2. in their article we obtain that the $L^2(\Omega)$ limit

$$D^u(\omega) := \lim_{n\to\infty} \sum_{t=0}^{n} \mathcal{P}_0(\langle X_t, u\rangle)e^{-\mathrm{i}t\omega} = \lim_{n\to\infty}\langle Z_n(\omega), u\rangle \tag{2.7}$$

exists on M_u and that $f^u(\omega) = \mathrm{Var}\Big(D^u(\omega)\Big)$. (Directly using their arguments would require to use the projection operator $\mathcal{P}_0^u(\cdot) = E[\,\cdot\,|\mathcal{G}_0(u)] - E[\,\cdot\,|\mathcal{G}_{-1}(u)]$.) We assume without loss of generality that $M = M_0 \cap M_1$ is a subset of M_u, otherwise replace M_u by $M_u \cap M$. We now determine $f^u(\omega)$. By result (P2) of Section 2.4.1 it follows that $E(\langle Z_{n_k}(\omega), u\rangle - \langle Z(\omega), u\rangle)^2 \to 0$ for every $u \in H$ and all $\omega \in M_1$. Hence, by result (P1) in the same section, we get

$$\mathrm{Var}\Big(\langle Z_{n_k}(\omega), u\rangle\Big) \to \langle \mathcal{F}_\omega(u), u\rangle < \infty,$$

for all $\omega \in M$ and all $u \in H$, which implies on M_u the relation $f^u(\omega) = \langle \mathcal{F}_\omega(u), u\rangle < \infty$. This shows part **(a)** on M_u.
We have $E[\langle X_t, u\rangle|\mathcal{G}_{-\infty}] = \langle E[X_t|\mathcal{G}_{-\infty}], u\rangle$ and by Assumption 1 this is equal to zero. The tower property of conditional expectations implies $E[\langle X_t, u\rangle|\mathcal{G}_{-\infty}(u)] = 0$ and hence on M_u **(b)** directly follows from Peligrad and Wu (2010). Let us note that their CLT result is stated for real valued time series, but this requirement is not needed. Hence we can apply it for the time series $(\langle X_t, u\rangle\colon t \geq 1)$ which takes values in $\mathbb{C}$ when $u \in H$.

It remains to prove that for all u in H we can find a common exceptional set of Lebesgue measure 0. To this end let H' be a dense and countable subset of H. We set $\widetilde{M} = \bigcap_{u' \in H'} M_u$. Then $(-\pi, \pi] \backslash \widetilde{M}$ has Lebesgue measure 0. Furthermore, for all $u' \in H'$ and $\omega \in \widetilde{M}$ **(a)** and **(b)** hold. The objective is now to extend this result to all $u \in H$. For **(a)** we observe that

$$
\begin{aligned}
&\Big|\langle \mathcal{F}_{n;\omega}(u), u\rangle - \langle \mathcal{F}_{\omega}(u), u\rangle\Big| \\
&\quad \leq \Big|\langle \mathcal{F}_{n;\omega}(u), u\rangle - \langle \mathcal{F}_{n;\omega}(u'), u'\rangle\Big| + \Big|\langle \mathcal{F}_{\omega}(u'), u'\rangle - \langle \mathcal{F}_{\omega}(u), u\rangle\Big| \\
&\qquad + \Big|\langle \mathcal{F}_{n;\omega}(u'), u'\rangle - \langle \mathcal{F}_{\omega}(u'), u'\rangle\Big| \\
&\quad \leq \Big[\operatorname{tr}\Big(\mathcal{F}_{n;\omega}\Big) + \operatorname{tr}\Big(\mathcal{F}_{\omega}\Big)\Big] \times \Big[\Big(\|u\| + \|u'\|\Big) \times \|u - u'\|\Big] \\
&\qquad + \Big|\langle \mathcal{F}_{n;\omega}(u'), u'\rangle - \langle \mathcal{F}_{\omega}(u'), u'\rangle\Big|.
\end{aligned}
$$

Since we can assume without loss of generality that $M_2 \subset \widetilde{M}$, it follows that for all $\omega \in \widetilde{M}$

$$
\limsup_{n\to\infty} \Big|\langle \mathcal{F}_{n;\omega}(u), u\rangle - \langle \mathcal{F}_{\omega}(u), u\rangle\Big| \leq 4\varepsilon(\|u\| + 1)\operatorname{tr}\Big(\mathcal{F}_{\omega}\Big),
$$

if $\|u - u'\| \leq \varepsilon \leq 1$. Since ε can be chosen arbitrarily small result **(a)** follows.
The proof of part **(b)** follows along similar lines of arguments. Just compare the characteristic functions of the real and complex part of $\langle \mathcal{X}_n(\omega), u\rangle$ to the corresponding normal ones. $\square$

2.4.2 Tightness

The following technical lemma will be crucial for showing tightness.

Lemma 6. *Consider sequences $(p_j^{(n)})_{j\geq 1}$, $n \geq 0$, with the following properties: (a) $p_j^{(n)} \geq 0$ for all j, n; (b) $\lim_n p_j^{(n)} = p_j^{(0)}$; (c) $\sum_{j\geq 1} p_j^{(0)} = p < \infty$; (d) $\lim_n \sum_{j\geq 1} p_j^{(n)} = p$; (e) $\sum_{j\geq 1} p_j^{(n)} < \infty$ for all $n \geq 1$. Then*

$$
\lim_{m\to\infty} \sup_n \sum_{j>m} p_j^{(n)} = 0.
$$

Proof. Fix an $\varepsilon > 0$. We have to show that for $m \geq m(\varepsilon)$ we have $\sum_{j\geq m} p_j^{(n)} < \varepsilon$ for all $n \geq 1$.
By (c) we can choose $m_1 = m_1(\varepsilon)$ such that $\sum_{j\geq m} p_j^{(0)} < \varepsilon/3$ for all $m \geq m_1$. Furthermore, by (b) we can choose a large enough $n_1 = n_1(\varepsilon)$ such that for all $n \geq n_1$ we have $|\sum_{j=1}^{m_1} p_j^{(0)} - \sum_{j=1}^{m_1} p_j^{(n)}| < \varepsilon/3$. Next, by possibly further enlarging n_1 we deduce from (c) and (d) that $|\sum_{j\geq 1} p_j^{(n)} - \sum_{j\geq 1} p_j^{(0)}| < \varepsilon/3$. Consequently, for $n \geq n_1$, we have

$$
\begin{aligned}
\sum_{j>m_1} p_j^{(n)} &= \sum_{j\geq 1} p_j^{(n)} - \sum_{j=1}^{m_1} p_j^{(n)} \\
&\leq |\sum_{j\geq 1} p_j^{(n)} - \sum_{j\geq 1} p_j^{(0)}| + |\sum_{j\geq 1} p_j^{(0)} - \sum_{j=1}^{m_1} p_j^{(0)}| + |\sum_{j=1}^{m_1} p_j^{(0)} - \sum_{j=1}^{m_1} p_j^{(n)}| \\
&< \varepsilon.
\end{aligned}
$$

Because of (a) this bound is still valid for all $m \geq m_1$. For the n_1 just chosen, we can find an $m_2 = m_2(\varepsilon)$, such that $\sum_{j>m_2} p_j^{(n)} < \varepsilon$ for all $n \leq n_1$. This is because of (d) and (e) we know that $\sup_{n\geq 1} \sum_{j\geq 1} p_j^{(n)} < \infty$. And again, because of (a) we know also that $\sum_{j>m} p_j^{(n)} < \varepsilon$ for all $m \geq m_2$ and $n \leq n_1$. Hence, set $m(\varepsilon) = \max\{m_1, m_2\}$. □

Lemma 7. *Take some ONB $(v_j\colon j \geq 1)$ of H. Lemma 6 applies with $p_j^{(n)} = \langle \mathcal{F}_{n;\omega}(v_j), v_j\rangle$, $p_j^{(0)} = \langle \mathcal{F}_{\omega}(v_j), v_j\rangle$ for all $\omega \in \widetilde{M}$.*

Proof. We can assume that $(v_j)_{j\geq 1}$ belongs to the dense subset H' which was defined in Proposition 1. Relation (a) is trivial. Relation (b) follows from part **(a)** of Proposition 1. Relation (c) holds because $\mathcal{F}_\omega$ is nuclear on $\widetilde{M}$. And similarly relation (e) holds because $\mathcal{F}_{n;\omega}$ is nuclear for any n. Finally note that (d) can be reformulated as $\operatorname{tr}\Big(\mathcal{F}_{n;\omega}\Big) \to \operatorname{tr}\Big(\mathcal{F}_{\omega}\Big)$. By Lemma 4 this holds for almost all $\omega \in M_2 \subset \widetilde{M}$. □

Lemma 8. *Under Assumptions 1 the sequence $(\mathcal{X}_n(\omega))_{n\geq 1}$ is tight for all $\omega \in \widetilde{M}$.*

Proof. Let $\varepsilon > 0$. We consider the sequences $0 < \ell_k \nearrow \infty$ and $0 < N_k \nearrow \infty$ and define

$$K = \bigcap_{k=1}^{\infty} \left\{ x \in H : \sum_{j>N_k} |\langle v_j, x\rangle|^2 \leq \frac{1}{\ell_k} \right\}.$$

Just as in Bosq (2000) (p. 52) we can see that it is a compact subset of H. We now have that

$$\begin{aligned} P\left(\mathcal{X}_n(\omega) \in K\right) &\geq 1 - \sum_{k=1}^{\infty} \ell_k \sum_{j>N_k} E\Big|\langle \mathcal{X}_n(\omega), v_j\rangle\Big|^2 \\ &= 1 - \sum_{k=1}^{\infty} \ell_k \sum_{j>N_k} \langle \mathcal{F}_{n;\omega}(v_j), v_j\rangle, \end{aligned}$$

where we used the σ-subadditivity and the Markov inequality. By Lemma 7 we know that

$$\sup_n \sum_{j\geq m} \langle \mathcal{F}_{n;\omega}(v_j), v_j\rangle \to 0 \quad (m \to \infty).$$

Therefore, for any $\varepsilon > 0$, we can choose increasing sequences (ℓ_k) and (N_k) such that

$$\ell_k \sum_{j>N_k} \langle \mathcal{F}_{n;\omega}(v_j), v_j\rangle \leq \varepsilon 2^{-k}.$$

□

2.4.3 Proofs of Lemma 1 and Theorems 1, 2 and 4

Proof of Lemma 1. It is easy to see that **(A3)** implies **(A1)** for all $\omega \in (-\pi, \pi]$. Now we prove **(A2)**. By part *(iii)* of Lemma 2 we have

$$E\|E[n^{1/2}\mathcal{X}_n(\omega)|\mathcal{G}_0]\|^2 = n\sum_{j\geq 0} E\|\mathcal{P}_{-j}(E[\mathcal{X}_n(\omega)|\mathcal{G}_0])\|^2 = n\sum_{j\geq 0} E\|\mathcal{P}_{-j}(\mathcal{X}_n(\omega))\|^2.$$

Therefore we may conclude that

$$\begin{aligned}
E\|E[n^{1/2}\mathcal{X}_n(\omega)|\mathcal{G}_0]\|^2 &\leq \sum_{j=0}^{\infty} E \sum_{s,t=1}^{n} |\langle \mathcal{P}_{-j}(X_t), \mathcal{P}_{-j}(X_s)\rangle| \\
&= \sum_{j=0}^{\infty}\sum_{s,t=1}^{n} E|\langle \mathcal{P}_{-j}(X_t), \mathcal{P}_{-j}(X_s)\rangle| \\
&\leq \sum_{s=1}^{n}\sum_{j=0}^{\infty}\left(\sum_{t=0}^{\infty}\nu_2(\mathcal{P}_0(X_{t+j}))\right)\nu_2(\mathcal{P}_0(X_{s+j})) = o(n).
\end{aligned}$$

With the help of Lemma 2 we obtain

$$\begin{aligned}
\|C_h\|_{\mathcal{S}} = \|EX_h \otimes X_0\|_{\mathcal{S}} &\leq \sum_{k=0}^{\infty} \|E\mathcal{P}_{-k}(X_h)\otimes \mathcal{P}_{-k}(X_0)\|_{\mathcal{S}} \\
&\leq \sum_{k=0}^{\infty} \nu_2\left(\mathcal{P}_0(X_{h+k})\right)\nu_2\left(\mathcal{P}_0(X_k)\right).
\end{aligned}$$

Consequently

$$\sum_{h=-\infty}^{\infty}\|C_h\|_{\mathcal{S}} \leq \left(\sum_{h=0}^{\infty}\nu_2\left(\mathcal{P}_0(X_h)\right)\right)^2 < \infty.$$

This proves *(ii)* and implies that expression (2.1) is well defined and continuous.
Since now we know that Assumption 2 holds, we infer from Theorem 2 that the Cèsaro means $\mathcal{F}_{n;\omega} \overset{\mathrm{w}}{\to} \mathcal{F}_\omega$, with $\mathcal{F}_\omega$ defined as in (2.4). But *(ii)* implies that also the regular partial sums $\sum_{|h|<n} C_h e^{-\mathrm{i}h\omega}$ converge to the same limit. □

Proof of Theorem 1. Parts **(II)** and **(III)** of Theorem 1 follow directly from Lemmas 4 and 5. Part **(I)** can be deduced from the polarization identity for self-adjoint operators Γ

$$\begin{aligned}
\langle \Gamma(x), y\rangle = \frac{1}{4}\Big[&\langle \Gamma(x+y), x+y\rangle - \langle \Gamma(x-y), x-y\rangle \\
&+ \mathrm{i}\langle \Gamma(x+\mathrm{i}y), x+\mathrm{i}y\rangle - \mathrm{i}\langle \Gamma(x-\mathrm{i}y), x-\mathrm{i}y\rangle\Big],
\end{aligned}$$

and from part **(a)** of Proposition 1. Next, the asymptotic normality of $\mathcal{X}_n(\omega)$ for all $\omega \in \widetilde{M}$ follows from the corresponding convergence of the projections (Proposition 1, part **(b)**) and the tightness shown in Lemma 8.
Finally, the asymptotic independence relation **(IV)** can be obtained by verifying that the projections $(\langle \mathcal{X}_n(\omega), u\rangle, \langle \mathcal{X}_n(\omega'), u'\rangle)$ converge for any u and u' in H to a bivariate complex Gaussian vector with independent components. Similarly as remarked by Peligrad and Wu (2010) this amounts to combining the scalar proof with a Wold argument. We sketch the main steps:
Define $\sigma_1^2 = \sigma_1^2(\omega, u) = \langle \mathcal{F}_\omega(u), u\rangle$ and $\sigma_2^2 = \sigma_2^2(\omega', u') = \langle \mathcal{F}_{\omega'}(u'), u'\rangle$ and assume below that $u, u' \in H$ and $\omega, \omega' \in \widetilde{M}$ (as defined in Proposition 1) with the additional constraints $\omega \neq \omega'$ and $\omega \neq -\omega'$ (which only exclude an additional null-set of $(-\pi, \pi]^2$). We will prove that for any such choice of u, u', ω, ω'

$$(\langle \mathcal{X}_n(\omega), u\rangle + \langle \mathcal{X}_n(\omega'), u'\rangle)\sqrt{n} \xrightarrow{d} \mathcal{CN}_{\mathbb{C}}(0, \sigma_1^2 + \sigma_2^2, c), \tag{2.8}$$

where $c := \Big(\langle \mathcal{F}_\omega(\bar{u}), u\rangle I\Big\{\omega \in \{0,\pi\}\Big\} + \langle \mathcal{F}_\omega(\bar{u}), u\rangle I\Big\{\omega' \in \{0,\pi\}\Big\}\Big)$. From this the asymptotic independence and joint Gaussianity can be deduced.
From Proposition 1 and its proof we know that $(\langle Z_n(\omega), u\rangle \colon n \geq 1)$ given in (2.7) converges in $L^2(\Omega)$ to some variable $D^u(\omega)$ with variance $\sigma_1^2(\omega, u) < \infty$. This provides the ingredient for the construction of the martingale approximation as given in (Peligrad and Wu, 2010, p. 11f.). Following their construction we obtain stationary and ergodic martingale difference sequences $(D_k^u(\omega) \colon k \geq 1)$ such that $D_k^u(\omega) \stackrel{d}{=} D^u(\omega)$ and such that

$$\langle n^{1/2}\mathcal{X}_n(\omega), u\rangle = \sum_{k=1}^n D_k^u(\omega)e^{-ik\omega} + R_n \quad \text{and} \quad \langle n^{1/2}\mathcal{X}_n(\omega'), u'\rangle = \sum_{k=1}^n D_k^{u'}(\omega')e^{-ik\omega'} + R_n',$$

where $E|R_n|^2 + E|R_n'|^2 = o(n)$. (See also the discussion after our Theorem 2.) For simplicity we write henceforth $D_k = D_k^{u'}(\omega')$ and $D_k' = D_k^{u'}(\omega')$ and set $Y_k := D_k e^{-ik\omega} + D_k' e^{-ik\omega'}$. Then $\sum_{k=1}^n Y_k$ is itself a martingale to which we will apply a martingale CLT. Due to

$$S_n := \langle n^{1/2}\mathcal{X}_n(\omega), u\rangle + n^{1/2}\mathcal{X}_n(\omega'), v\rangle = \sum_{k=1}^n Y_k + R_n + R_n'$$

this CLT carries over to S_n. The Lindeberg condition for this martingale is easily checked. As for the asymptotic *variance* we have

$$\begin{aligned}
\frac{1}{n}\sum_{k=1}^n E[|Y_k|^2\,|\mathcal{G}_{k-1}] &= \frac{1}{n}\sum_{k=1}^n E[|D_k|^2\,|\mathcal{G}_{k-1}] + \frac{1}{n}\sum_{k=1}^n E[D_k\overline{D_k'}\,|\mathcal{G}_{k-1}]e^{-ik(\omega-\omega')} \\
&\quad + \frac{1}{n}\sum_{k=1}^n E[D_k'\overline{D_k}\,|\mathcal{G}_{k-1}]e^{-ik(\theta'-\theta)} + \frac{1}{n}\sum_{k=1}^n E[|D_k'|^2\,|\mathcal{G}_{k-1}] \\
&\xrightarrow{a.s.} \quad E|D_1|^2 + E|D_1'|^2 = \sigma_1^2 + \sigma_2^2,
\end{aligned}$$

where we applied the ergodic theorem and Lemma 5 in Wu (2004) for the middle terms. This Lemma applies if $\omega \neq \omega'$. If $\omega + \omega' \neq 0$ we obtain by the same lemma for the asymptotic *relation*

$$\begin{aligned}
\frac{1}{n}\sum_{k=1}^n E[Y_k^2\,|\mathcal{G}_{k-1}] &= \frac{1}{n}\sum_{k=1}^n E[D_k^2\,|\mathcal{G}_{k-1}]e^{-2ik\omega} \\
&\quad + \frac{1}{n}\sum_{k=1}^n E[2D_kD_k'e^{-ik(\omega+\omega')}\,|\mathcal{G}_{k-1}] + \frac{1}{n}\sum_{k=1}^n E[(D_k')^2\,|\mathcal{G}_{k-1}]e^{-2ik\omega'} \\
&\xrightarrow{a.s.} \quad ED_1^2 I\{\omega \in \{0,\pi\}\} + E(D_1')^2 I\{\omega \in \{0,\pi\}\} = c.
\end{aligned}$$

□

Proof of Theorem 2. Let $\omega_0 \in (-\pi, \pi]$ be such that Assumption 2 is satisfied. For notational convenience we use for the proof $\omega = \omega_0$. Due to relation (2.3) and stationarity, we have that

$$\begin{aligned}
E\|\mathcal{X}_n(\omega)\|^2 &= \frac{1}{n}\sum_{k=1}^n E\|Z_{n-k}^{(k)}(\omega)\|^2 + E\Big\|E[\mathcal{X}_n(\omega)|\mathcal{G}_0]\Big\|^2 \\
&= E\|Z^{(1)}(\omega)\|^2 + o(1),
\end{aligned}$$

hence **(II)** holds. Moreover, we have

$$E|\langle \mathcal{X}_n(\omega), u\rangle|^2 = E|\langle Z^{(1)}(\omega), u\rangle|^2 \ + \ o(1)$$

and with the polarization identity this shows **(I)**. The CLT for the projections $\langle \mathcal{X}_n(\omega), u\rangle$, as well as its bivariate version when considering two frequencies, can be shown by the same martingale approximation as in the proof of Theorem 1. Also the proof of tightness of $\mathcal{X}_n(\omega)$ doesn't require any new ideas and can be shown along the same lines as in the proof of Theorem 1. □

Proof of Theorem 4. The first step is again to show that $\mathcal{X}_n(\omega_n)$ can be approximated by a martingale just as in the case when the frequency is fixed. From Lemma 1 we know that both conditions of Assumption 2 are satisfied for any frequency. We may thus define $Z^{(k)}(\omega) := \sum_{t=0}^{\infty} \mathcal{P}_k(X_{k+t})e^{-\mathrm{i}t\omega}$ (as a limit in $L^2_H(\Omega)$) and the variables $\widetilde{\mathcal{X}}_n(\omega_n) := n^{-1/2}\sum_{k=1}^{n} Z^{(k)}(\omega)e^{-\mathrm{i}k\omega_n}$. Then

$$n^{1/2}\left(\mathcal{X}_n(\omega_n) - \widetilde{\mathcal{X}}_n(\omega_n)\right) = \sum_{k=1}^{n}\left(\mathcal{P}_k(S_n(\omega_n)) - Z^{(k)}(\omega)e^{-\mathrm{i}k\omega_n}\right) + E[n^{1/2}\mathcal{X}_n(\omega_n)|\mathcal{G}_0].$$

From the proof of Lemma 1 we infer that $\|E[n^{1/2}\mathcal{X}_n(\omega_n)|\mathcal{G}_0]\| = o(n)$. Furthermore, using orthogonality, we get

$$\begin{aligned}
&\nu_2^2\left(\sum_{k=1}^{n}\left(\mathcal{P}_k(n^{1/2}\mathcal{X}_n(\omega_n)) - Z^{(k)}(\omega)e^{-\mathrm{i}k\omega_n}\right)\right)\\
&\quad= \sum_{k=1}^{n}\nu_2^2\left(\sum_{t=k}^{n}\mathcal{P}_k(X_t)e^{-\mathrm{i}t\omega_n} - Z^{(k)}(\omega)e^{-\mathrm{i}k\omega_n}\right)\\
&\quad= \sum_{k=1}^{n}\nu_2^2\left(Z^{(k)}_{n-k}(\omega_n) - Z^{(k)}(\omega)\right).
\end{aligned}$$

It is easy to show that **(A3)** and $\omega_n \to \omega$ imply that the last term is $o(n)$.

Fix some arbitrary $u, u' \in H$, define σ_1^2 and σ_2^2 as in the proof of Theorem 1. Moreover, we set $D_k := \lim_{n\to\infty}\langle Z^{(k)}_n(\omega_n), u\rangle$ and $D'_k = \lim_{n\to\infty}\langle Z^{(k)}_n(\omega'_n), u'\rangle$ and $Y^n_k := D_k e^{-\mathrm{i}k\omega_n} + D'_k e^{-\mathrm{i}k\omega'_n}$. We show that

$$(\langle \widetilde{\mathcal{X}}_n(\omega), u\rangle + \langle \widetilde{\mathcal{X}}_n(\omega'), u'\rangle) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} Y^n_k \xrightarrow{d} \mathcal{CN}_{\mathbb{C}}(0, \sigma_1^2 + \sigma_2^2, 0).$$

To this end we apply a martingale CLT to the array of martingale differences $(Y^n_k)_{1\le k\le n}$, $n \ge 1$. The Lindeberg condition is easily checked and it remains again to determine the asymptotic *variance* and *relation*. In analogy to the proof of Theorem 1 we get

$$\begin{aligned}
\frac{1}{n}\sum_{k=1}^{n} E[|Y^n_k|^2\,|\mathcal{G}_{k-1}] &= \frac{1}{n}\sum_{k=1}^{n} E[|D_k|^2\,|\mathcal{G}_{k-1}] + \frac{1}{n}\sum_{k=1}^{n} E[D_k\overline{D'_k}\,|\mathcal{G}_{k-1}]e^{-\mathrm{i}k(\omega_n-\omega'_n)}\\
&\quad+ \frac{1}{n}\sum_{k=1}^{n} E[D'_k\overline{D_k}\,|\mathcal{G}_{k-1}]e^{-\mathrm{i}k(\omega'_n-\omega_n)} + \frac{1}{n}\sum_{k=1}^{n} E[|D'_k|^2\,|\mathcal{G}_{k-1}],
\end{aligned}$$

and by the exact same arguments the first and the last term converge to σ_1^2 and σ_2^2, respectively. In order to deal with the middle term we set $U_k := E[D_k\overline{D_k'}|\mathcal{G}_{k-1}]$ and our target is to prove that

$$\frac{1}{n}\sum_{k=1}^{n} U_k e^{-\mathrm{i}k(\omega_n-\omega_n')} \xrightarrow[n\to\infty]{P} 0.$$

The third term can be treated analogously. For some $m \geq 1$ set $U_k^m = E[U_k|\varepsilon_{k-1},\ldots,\varepsilon_{k-m-1}]$. Then $(U_k^m)_{k\geq 1}$ is a strictly stationary and m-dependent sequence which satisfies $U_k^m \xrightarrow{L^1} E[U_k|\varepsilon_{k-1},\ldots] = U_k$ $(m\to\infty)$. Thus, for any $\epsilon > 0$ we can find a large enough $m \geq 1$ such that $E|U_1 - U_1^m| \leq \epsilon^2$. Now we set $\alpha_n := \omega_n - \omega_n'$ and note that by our assumptions $\sum_{k=1}^n e^{-\mathrm{i}k\alpha_n} = 0$. It follows that

$$\begin{aligned}
&P\left(\left|\frac{1}{n}\sum_{k=1}^{n} U_k e^{-\mathrm{i}k\alpha_n}\right| > \epsilon\right)\\
&\quad\leq P\left(\left|\frac{1}{n}\sum_{k=1}^{n}(U_k - U_k^m)e^{-\mathrm{i}k\alpha_n}\right| > \epsilon/2\right) + P\left(\left|\frac{1}{n}\sum_{k=1}^{n} U_k^m e^{-\mathrm{i}k\alpha_n}\right| > \epsilon/2\right)\\
&\quad\leq \frac{\mathrm{E}\,|U_1 - U_1^m|}{\epsilon/2} + P\left(\left|\frac{1}{n}\sum_{j=1}^{m}\sum_{\substack{1\leq k\leq n\\ m|(k-j)}}\Big(U_k^m - EU_k^m\Big)e^{-\mathrm{i}k\alpha_n}\right| > \epsilon/2\right)\\
&\quad\leq \epsilon/2 + \sum_{j=1}^{m} P\left(\left|\frac{1}{n/m}\sum_{\substack{1\leq k\leq n\\ m|(k-j)}}\Big(U_k^m - EU_k^m\Big)e^{-\mathrm{i}k\alpha_n}\right| > \epsilon/2\right). \qquad (2.9)
\end{aligned}$$

Here $m|\ell$ signifies that m is a divisor of ℓ. For $k = j, j+m, j+2m, \ldots$ the terms U_k^m are i.i.d. Thus, we can apply the weighted law of large numbers in Pruitt (1966) to obtain that for large enough $n \geq n_0(\epsilon, m)$ each of the probabilities in (2.9) is $\leq \epsilon/(2m)$.
By the same kind of arguments it can be seen that the asymptotic relation is zero.
Finally, in order to show the tightness of $\mathcal{X}_n(\omega_n)$, we need to apply Lemma 6 with $p_j^{(n)} = E|\langle\mathcal{X}_n(\omega_n), v_j\rangle|^2$, and $p_j^{(0)} = E|\langle\mathcal{F}_\omega(v_j), v_j\rangle| = E|\langle Z^{(1)}(\omega\omega), v_j\rangle|^2$, where $(v_j)_{j\geq 1}$ is an arbitrary ONB. Using the previous approximation $\mathcal{X}_n(\omega_n) = \widetilde{\mathcal{X}}_n(\omega) + o_{L^2}(1)$, it is easy to check that conditions (a)-(e) are satisfied. With this we can conclude along the lines of Section 2.4.2. □

3 On the maximal norm of the functional periodogram

Clément Cerovecki*, Vaidotas Characiejus† and Siegfried Hörmann‡

Keywords: Functional time series, Gaussian approximation, Extreme value distribution, Fourier transform

Abstract

We consider the periodgram of multivariate and functional data and derive the limiting distribution of the maximal norm over fundamental frequencies. We provide conditions which assure that this maximum is in the domain of attraction of the Gumbel distribution. Our results generalize a theorem of Davis and Mikosch (1999) to multivariate and functional data. We consider an application to testing for hidden periodic patterns in functional time series.

3.1 Introduction

Consider some time series $(X_t)_{t\in\mathbb{Z}}$. Let $\mathrm{i} = \sqrt{-1}$ and let

$$\mathcal{X}_n(\omega) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t e^{-\mathrm{i}t\omega}$$

be the *discrete Fourier transform (DFT)* of the sample. Define

$$I_n(\omega_j) := |\mathcal{X}_n(\omega_j)|^2, \quad \omega_j = \frac{2\pi j}{n}, \quad 1 \le j \le q := \left\lfloor \frac{n-1}{2} \right\rfloor. \tag{3.1}$$

Here $|z|^2 = z\bar{z}$ is the squared modulus of some complex number z. Then $I_n(\omega_j)$ are the *periodogram ordinates* of the sample $X_1, \dots, X_n$ on fundamental frequencies ω_j. The periodogram is a well known and important tool in time series analysis. It is the main ingredient for estimation of the spectral density $f(\omega) := \sum_{k\in\mathbb{Z}} \mathrm{Cov}(X_k, X_0)e^{-\mathrm{i}k\omega}$ of the series $(X_t)_{t\in\mathbb{Z}}$ as well as the key statistic for detection of periodic signals in the data. For example, if

$$X_t = \mu_0 + \alpha\cos(\omega_j t + \vartheta) + \xi_t,$$

and if $\xi_1, \dots, \xi_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$, then $I_n(\omega_j)$ will diverge at rate n if $\alpha \neq 0$, while the $I_n(\omega_k)$, $k \neq j$, will be i.i.d. exponential variables. See e.g. Brockwell and Davis (1991).

In the above situation, when ω_j is an unknown frequency, it is natural to consider the statistic $M_n = \max_{j=1,\dots,q} I_n(\omega_j)$. Since under the null ($\alpha = 0$) the $I_n(\omega_j)$, $1 \le j \le q$, are

*Université libre de Bruxelles, Boulevard du Triomphe, CP210 B-1050 Brussels - Belgium Office: NO O.9.211. *E-Mail address*: clement.cerovecki@ulb.ac.be

†Université libre de Bruxelles, Boulevard du Triomphe, CP210 B-1050 Brussels - Belgium Office: NO O.9.211. *E-Mail address*: vaidotas.characiejus@ulb.ac.be

‡Université libre de Bruxelles, Boulevard du Triomphe, CP210 B-1050 Brussels - Belgium Office: NO O.9.115 *shormann@ulb.ac.be*

i.i.d. exponential variables, it follows immediately that M_n is in the domain of attraction of a Gumbel distribution. A natural question is what happens if the noise variables $(\xi_t)_{t\in\mathbb{Z}}$ are no longer i.i.d. Gaussian. To this end, consider $0<\lambda_1<\cdots<\lambda_m<\pi$ and fundamental frequencies $\omega_{\ell,n}\to\lambda_\ell$ $(n\to\infty)$. It is well known that for stationary time series it holds under certain regularity assumptions that

$$(I_n(\omega_{1,n}),\ldots,I_n(\omega_{m,n}))\xrightarrow{\mathrm{d}}(f(\lambda_1)E_1,\ldots,f(\lambda_m)E_m), \tag{3.2}$$

where $E_k\overset{\text{i.i.d.}}{\sim}\mathrm{Exp}(1)$. Such a result has e.g. been obtained by Peligrad and Wu (2010) under very sharp conditions. It suggests that for i.i.d. $(\xi_t)_{t\in\mathbb{Z}}$ (in this case $f(\lambda)\equiv\mathrm{Var}(\xi_1)$) the periodogram ordinates $(I_n(\omega_j))_{j=1,\ldots,q}$ behave approximately like i.i.d. exponential variables. Indeed, Davis and Mikosch (1999) have proven that the asymptotic Gumbel distribution of M_n still holds if $(\xi_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence with $s>2$ moments. This result has later been extended by Lin and Liu (2009) to a broad class of stationary processes. However, the problem is quite delicate and it is important to note that (3.2) cannot be directly applied since the statistic M_n depends on all fundamental frequencies, while the result in (3.2) applies to a fixed and finite number of frequencies.

The purpose of this article is to extend the result of Davis and Mikosch (1999) to high dimensional data. We are primarily interested in functional data, i.e. when each variable X_t takes values in some infinite dimensional function space H_0. In passing, we will also consider $H_0=\mathbb{R}^p$, $p\geq 1$. For the functional data, we restrict our attention to observations with values in some separable Hilbert space, like the space of square integrable functions on $[0,1]$: $H_0=L^2([0,1])$. Since the p-dimensional Euclidian space is also a separable Hilbert-space we usually don't have to distinguish between functional and multivariate in terms of notation. For example, the definition and notion of the DFT for functional data is the same as above in the scalar setting. Before we go more into details of our theory, let us first introduce some further notation which will be used throughout this paper.

By its very definition $\mathcal{X}_n(\omega)$ is an element in the complex Hilbert space $H:=H_0\oplus \mathrm{i}H_0$. The space H_0 is equipped with some inner product $\langle\cdot,\cdot\rangle$ and corresponding norm $\|\cdot\|$. The space H inherits the Hilbert space structure from H_0. The complex inner product is defined as $\langle u,v\rangle_{H_0}=\langle u_0,v_0\rangle+\langle u_1,v_1\rangle+\mathrm{i}(\langle u_1,v_0\rangle-\langle u_0,v_1\rangle)$ for any $u=u_0+\mathrm{i}u_1$ and $v=v_0+\mathrm{i}v_1$ in H. For a lighter notation we will henceforth consider H_0 as a subspace of H and then use $\langle\cdot,\cdot\rangle$ for the real and the complex inner product. We do the same for the norm and other definitions to come. We denote by $\mathcal{L}(H)$ the space of bounded linear operators on H and equip it with the usual operator norm $\|\boldsymbol{\alpha}\|=\sup_{\|x\|=1}\|\boldsymbol{\alpha}(x)\|$. We say that an operator $\boldsymbol{\alpha}$ is Hilbert-Schmidt (trace-class) if its singular values $(\sigma_k)_{k\geq1}$ are square summable (absolutely summable). We define the corresponding Hilbert-Schmidt norm $\|\boldsymbol{\alpha}\|_{\mathcal{S}}=\left(\sum_{k=1}^\infty\sigma_k^2\right)^{1/2}$ and trace-class norm $\|\boldsymbol{\alpha}\|_{\mathcal{T}}=\sum_{k=1}^\infty\sigma_k$. We further define the outer-product operators $x\otimes y:\ u\mapsto x\langle u,y\rangle$, for $x,y\in H$. In particular this gives rise to the covariance operator $\mathrm{Var}(X):=E[(X-EX)\otimes(X-EX)]$. We note that for $H_0=\mathbb{R}^p$ this is the usual covariance matrix. For more details on random elements in Hilbert spaces, we refer to Bosq (2000).

With this notation at hand, we can define the functional periodogram:

$$I_n(\omega_k):=\mathcal{X}_n(\omega_k)\otimes\mathcal{X}_n(\omega_k),\quad \omega_k=\frac{2\pi k}{n},\quad 1\leq k\leq q:=\left\lfloor\frac{n-1}{2}\right\rfloor.$$

We refer to Panaretos and Tavakoli (2013a). In this article, we investigate the following statistic

$$M_n = \max_{j=1,\dots,q} \|I_n(\omega_j)\| = \max_{j=1,\dots,q} \|\mathcal{X}_n(\omega_j)\|^2. \tag{3.3}$$

The main objective of this paper is to obtain the limiting distribution of M_n when $X_1, X_2, \dots$ is a functional random sample. To this end we first tackle the multivariate case and show that if $E\|X_1\|^s < \infty$ for some $s > 2$ then M_n is in the domain of attraction of the Gumbel distribution. To the best of our knowledge this result is the first multivariate extension of Davis and Mikosch (1999). We also provide a further extension to linear processes following the classical approach of Walker (1965). As an application we show how to extend some of the tests for hidden periodic signals proposed by Hörmann et al. (2017) to the case where the frequency is unknown.

In a second step we pass from multivariate to infinite dimensional data. We follow a common approach in FDA, which is to project the functional data on the first $p = p_n$ eigenfunctions (principal components) of $\mathrm{Var}(X_1)$ and let $p \to \infty$. This step is technically quite delicate. The main ingredient of our proof is based on a Gaussian approximation result from Chernozhukov et al. (2017) which in turn relies on a powerful anti-concentration inequality due to Nazarov (2003).

The rest of the paper is organized as follows, in Section 3.2 we formulate our main theorems. We illustrate in Section 3.3 how to use these results for detecting a periodic signal in functional time series at some unknown frequency. The proofs are given in Section 3.4.

3.2 Main results

In the following we assume that $X_1, X_2, \dots$ are i.i.d. functional data with covariance operator $C = \mathrm{Var}(X_1)$. We denote by $v_1, v_2, \dots$ the eigenfunctions of C with corresponding eigenvalues $\lambda_1 > \lambda_2 > \cdots$. To simplify some arguments we assume throughout distinct eigenspaces, which is a common assumption in functional data analysis. For some integer $p \geq 1$ we further define

$$\mathcal{X}_n^p(\omega) = \sum_{k=1}^{p} \sum_{t=1}^{n} \frac{1}{\sqrt{n}} \langle X_t, v_k \rangle v_k e^{-\mathrm{i}t\omega} \quad \text{and} \quad I_n^p(\omega) = \mathcal{X}_n^p(\omega) \otimes \mathcal{X}_n^p(\omega). \tag{3.4}$$

In other words we consider $\mathcal{X}_n(\omega)$ and $I_n(\omega)$ as defined in Section 3.1, but with X_t replaced by $\mathcal{P}_{\{\mathrm{span}(v_1,\dots,v_p)\}}(X_t)$, i.e. the projection of X_t onto the space spanned by the first p eigenfunctions of C. Obviously, if $X_t \in \mathbb{R}^p$ then $\mathcal{X}_n^p(\omega) = \mathcal{X}_n(\omega)$.

We will now investigate the asymptotic behaviour of $M_n^p = \max_{1 \leq j \leq q} \|\mathcal{X}_n^p(\omega_j)\|^2$ when p is fixed. Later we will consider the case $p \to \infty$. To this end let us introduce the following centering constants:

$$b_n^p = \lambda_1 \log(n\alpha_{1,p}/2), \quad \text{where} \quad \alpha_{1,p} = \prod_{j=2}^{p} (1 - \lambda_j/\lambda_1)^{-1}. \tag{3.5}$$

Theorem 8. *If $E\|X_1\|^s < \infty$, for some $s > 2$, we have that*

$$\frac{M_n^p - b_n^p}{\lambda_1} \xrightarrow{d} \mathcal{G} \quad (n \to \infty), \tag{3.6}$$

where $\mathcal{G}$ follows a standard Gumbel distribution.

The proof of Theorem 8 mainly relies on a Gaussian approximation due to Chernozhukov et al. (2017). These authors provide high-dimensional uniform Berry-Esseen-type bounds for the multivariate central limit theorem. We summarize this result and provide the explicit bounds which we need in Section 3.4.3. After this Gaussian approximation step, it will turn out that we can restrict our asymptotic investigation to that of $\widetilde{M}_n^p$, where $\widetilde{M}_n^p$ is just defined as M_n^p but with X_t replaced by some i.i.d. Gaussian sequence $(Z_t)_{t\in\mathbb{Z}}$ with $\mathrm{Var}(Z_t) = \mathrm{Var}(X_t)$.

As it was already hinted above, Theorem 8 applies to the particular case when H is the p-dimensional Euclidean space. To our knowledge, this is the first multivariate generalization of Theorem 2.1 in Davis and Mikosch (1999). In Section 3.3.1 we will further investigate this finite dimensional setting, in particular we will extend it to non i.i.d. processes.

We now present a first extension of Theorem 8, when $p = p_n \nearrow \infty$.

Theorem 9. *We assume that $E\|X_1\|^4 < \infty$, and that there exists an $k_0 \geq 2$ such that*

$$k\lambda_k \geq (k+1)\lambda_{k+1}, \quad \text{for all} \quad k \geq k_0. \tag{3.7}$$

Then the conclusion of Theorem 8 still holds if p is replaced by an increasing sequence of integers $(p_n)_{n\geq 1}$ which satisfies conditions

$$\frac{p_n^3}{\lambda_{p_n}^{1/2}} = o\left(\frac{n^{1/6}}{\log^{7/6}(n)}\right) \qquad \text{and} \qquad p_n = O(n^{\gamma_0}), \tag{3.8}$$

with

$$\gamma_0 \ < \ \min\left\{\min_{2\leq k\leq k_0} \frac{1}{k}\left(\frac{\lambda_1}{\lambda_k} - 1\right), 1\right\}. \tag{3.9}$$

Remark 1. Condition (3.7) is a very mild restriction. Indeed, since the eigenvalues are positive, decreasing and summable, we already know that $k\lambda_k \to 0$, when $k \to +\infty$. The condition further requires that the sequence $k\lambda_k$ is eventually decreasing.

If we let p grow to infinity, then we can choose a location sequence $(b_n)_{n\geq 1}$ in (3.6) which is independent of p. This is implied by the following simple lemma.

Lemma 9. *For any sequence of integers $p_n \nearrow \infty$, we have that*

$$|b_n^{p_n} - b_n| \longrightarrow 0 \quad (n \to \infty), \tag{3.10}$$

where $b_n = \lambda_1 \log(n/2) - \lambda_1 \sum_{j=2}^{\infty} \log(1 - \lambda_j/\lambda_1)$. The latter series converges.

We now present our last main result which provides the asymptotic behaviour of the *fully functional* statistic M_n. Not surprisingly, the necessary technical conditions are intimately connected with the decay rate of the eigenvalues $(\lambda_k)_{k\geq 1}$ of C.

Theorem 10. *We suppose that $E\|X_1\|^4 < \infty$, and that there is a sequence $(p_n)_{n\geq 1}$ which satisfies conditions from Theorem 9. Consider some sequence $(\ell_k)_{k\geq 1}$ of positive numbers such that $\sum_{k=1}^{\infty} \ell_k = 1$, some $s > 2$, and assume that*

$$\sum_{k>p_n} \ell_k^{-s/2} E\left|\langle X_1, v_k\rangle\right|^s = o(n^{s/2-1}) \tag{3.11}$$

and that

$$\sum_{k>p_n} (\lambda_k/\ell_k)^{s/2} = o(1/n), \tag{3.12}$$

Then $\lambda_1^{-1}(M_n - b_n) \xrightarrow{d} \mathcal{G}$, *where* $\mathcal{G}$ *follows a standard Gumbel distribution.*

Note that since $s/2 - 1 > 0$, a sufficient condition for (3.11) to hold is that

$$\sup_{k\geq 1} \frac{E|\langle X_1, v_k\rangle|^s}{\lambda_k^{s/2}} = C < \infty. \tag{3.13}$$

For example, this condition is satisfied by Gaussian processes for arbitrarily large s.

For the sake of explicit conditions we will consider in the corollary below two special cases. We let $a_k \sim b_k$ indicate that $\limsup |a_k/b_k| < \infty$ and $\limsup |b_k/a_k| < \infty$.

Corollary 1. *Suppose that* (3.7)*. Furthermore, assume that one of the conditions*

$$\lambda_k \sim \rho^k \quad \textit{for some } 0 < \rho < 1, \tag{\textbf{E1}}$$

$$\lambda_k \sim k^{-\nu} \quad \textit{for some } \nu > 1, \tag{\textbf{E2}}$$

is satisfied. Then under (**E1**) *conditions* (3.8), (3.11) *and* (3.12) *hold if* $E\|X\|^{2(s-1)} < \infty$ *with* $s > 6$*. Moreover, these conditions hold under* (**E2**) *if* ν *and* s *are such that* $s > \frac{2}{\nu-1}$ *and*

$$\frac{1}{(\nu-1)s/2-1} < \min\left\{\frac{s-2}{2+s-\nu}, \frac{1}{6\,(\nu/2+3)}, \min_{2\leq k\leq k_0} \frac{1}{k}\left(\frac{\lambda_1}{\lambda_k} - 1\right), 1\right\}. \tag{3.14}$$

Consequently, Theorem 10 applies.

We remark that for big enough s, (3.14) can always be verified.

So far we have assumed that C and its eigenvalues $(\lambda_k)_{k\geq 1}$ where known. In the next lemma we show that our results remain valid if we replace λ_1 and b_n^p by estimators $\widehat{\lambda}_1$ and $\widehat{b}_n^p = \widehat{\lambda}_1 \log(n\widehat{\alpha}_{1,p}/2)$, where $\widehat{\alpha}_{1,p} = \prod_{j=2}^p (1 - \widehat{\lambda}_j/\widehat{\lambda}_1)^{-1}$.

Lemma 10. *Let* $X_1, X_2, \ldots$ *be an i.i.d. sequence with* $E\|X_1\|^4 < \infty$*. Let* $\widehat{\lambda}_k$ *be the empirical eigenvalues, i.e. the eigenvalues of*

$$\widehat{C} = \frac{1}{n}\sum_{t=1}^n (X_t - \bar{X}) \otimes (X_t - \bar{X}).$$

If $p = o(n^{1/2})$ *and if* $M_n^p - b_n^p$ *converges weakly, then*

$$\lambda_1^{-1}(M_n^p - b_n^p) - \hat{\lambda}_1^{-1}(M_n^p - \widehat{b}_n^p) \xrightarrow{P} 0 \quad (n \to \infty). \tag{3.15}$$

3.3 Testing for hidden periodicities

In this section we present some statistical applications of Theorems 8 and 9. Similarly as in Hörmann et al. (2017) our objective is to test the presence of a periodic component in a functional times series. In contrast to the aforementioned authors, we do not assume that the frequency of the potential periodic signal is known. Our model is given as follows.

$$X_t = \mu + m_t + \varepsilon_t \tag{3.16}$$

where $\mu \in H$, $m : \mathbb{Z} \to H$ is a d–periodic deterministic function for some *unknown* integer $d \geq 1$, and $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a centered H-valued noise process. We impose the identifiability constraint $\sum_{t=1}^d m_t = 0$ and then consider the following testing problem:

$$\begin{cases} \mathcal{H}_0 : & m_1 = \cdots = m_d = 0, \\ \mathcal{H}_1 : & \max_{1\leq t\leq d} \|m_t\| > 0. \end{cases} \tag{3.17}$$

For the sake clarity we specify in this section the process on which our statistics are defined in terms of superscript, e.g. $\mathcal{X}_n^X(\omega) = n^{-1/2}\sum_{t=1}^n X_t e^{-\mathrm{i}t\omega}$ or $\mathcal{X}_n^\varepsilon(\omega) = n^{-1/2}\sum_{t=1}^n \varepsilon_t e^{-\mathrm{i}t\omega}$. We have that

$$M_n^{X,p} \geq \|\mathcal{X}_n^{X,p}(2\pi/d)\|^2 \geq \|\mathcal{X}_n^{m,p}(2\pi/d)\|^2 - 2\|\mathcal{X}_n^{\varepsilon,p}(2\pi/d)\| \cdot \|\mathcal{X}_n^{m,p}(2\pi/d)\|.$$

Furthermore, we know that $\|\mathcal{X}_n^{\varepsilon,p}(2\pi/d)\| \xrightarrow{d} \sqrt{U}$, where U is a hypo–exponential random variable whose parameters are the p first eigenvalues of $\mathcal{F}_{2\pi/d}$. Under the alternative we have for any $n > d$, that

$$\begin{aligned} x_n^2 := \|\mathcal{X}_n^{m,p}(2\pi/d)\|^2 &= \frac{1}{n}\sum_{k=1}^p \left|\sum_{t=1}^n \langle m_t, v_k\rangle e^{-\frac{\mathrm{i}t2\pi}{d}}\right|^2 \\ &\geq \frac{1}{n}\left\lfloor\frac{n}{d}\right\rfloor^2 \sum_{k=1}^p \left|\sum_{t=0}^{d-1}\langle m_t, v_k\rangle e^{-\frac{\mathrm{i}t2\pi}{d}}\right|^2 - \frac{1}{n}\sum_{k=1}^p \left|\sum_{t=n-d\lfloor n/d\rfloor+1}^n \langle m_t, v_k\rangle e^{-\frac{\mathrm{i}t2\pi}{d}}\right|^2. \end{aligned}$$

This term tends to ∞ at rate proportional to n whenever there is at least one $j \in \{1,\ldots,k\}$ such that $(\langle m_1, v_j\rangle, \ldots, \langle m_d, v_j\rangle)'$ is not orthogonal $(e^{-\frac{\mathrm{i}2\pi}{d}}, e^{-\frac{\mathrm{i}4\pi}{d}}, \ldots, e^{-\frac{\mathrm{i}2d\pi}{d}})'$. Therefore, we get that $\|\mathcal{X}_n^{\varepsilon,p}(2\pi/d)\|/x_n = o_P(1)$, which shows that under the alternative $M_n^p \xrightarrow{P} \infty$.

We will first assume that

Assumption 5. *The noise $(\varepsilon_t)_{t\in\mathbb{Z}}$ is i.i.d. with $E\|\varepsilon_1\|^4 < \infty$.*

Remark 2. From Proposition 10 and Assumption 5 we know that all the results of this section are valid if we replace a_n and b_n^p by their estimates. Moreover, the identifiability constraints ensures that the estimator are still convergent under the alternative, e.g. we have that

$$\overline{X} = \mu + \overline{\varepsilon} + \frac{1}{n}\sum_{t=n-d\lfloor n/d\rfloor+1}^n m_t \underset{n\to\infty}{\longrightarrow} \mu,$$

almost surely and in $L^1(\Omega)$.

3.3.1 Projection approach

In practise functional data are only available through a finite dimensional representation based on some appropriate smoothing method. It is thus natural to assume in a first step, that there exists a set of orthonormal functions $\varphi_1, \ldots, \varphi_p$ in H such that $X_t = \sum_{k=1}^{p} \langle X_t, \varphi_k \rangle \varphi_k$ for all $t \in \mathbb{Z}$. Let $\mathbf{X}_t = (\langle X_t, \varphi_1 \rangle, \ldots, \langle X_t, \varphi_p \rangle)'$, we consider the following multivariate analogue of model (3.16)

$$\mathbf{X}_t = \boldsymbol{\mu} + \mathbf{m}_t + \boldsymbol{\varepsilon}_t \tag{3.18}$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$, $\mathbf{m} : \mathbb{Z} \to \mathbb{R}^p$ is a d–periodic deterministic function for some integer $d \geq 1$, and $(\boldsymbol{\varepsilon}_t)_{t \in \mathbb{Z}}$ is a centered $\mathbb{R}^p$-valued process.

The following result is a direct consequence of Theorem 8:

Test 1. *Under Assumption 5, the test that rejects $\mathcal{H}_0$ if*

$$\hat{\lambda}_1^{-1}(M_n^{\mathbf{X}} - \hat{b}_n^p) \; > \; Q_{1-\alpha}^{\mathcal{G}}, \tag{3.19}$$

has asymptotic level α.

Here $Q_{1-\alpha}^{\mathcal{G}}$ is the $(1-\alpha)$–quantile of the Gumbel distribution.

Assuming i.i.d. innovations is certainly a restriction. We thus consider the following generalization:

Assumption 6. *The noise $(\boldsymbol{\varepsilon}_t)_{t \in \mathbb{Z}}$ is a linear process i.e. $\boldsymbol{\varepsilon}_t = \sum_{k \geq 0} \Psi_k \mathbf{Z}_{t-k}$ where $(\mathbf{Z}_t)_{t \in \mathbb{Z}}$ are i.i.d. innovations in $\mathbb{R}^p$ with $E\|\mathbf{Z}_1\|^4 < \infty$, $\mathrm{Var}(\mathbf{Z}_1) = \Sigma$ and*

$$\sum_{k=0}^{\infty} k^{1/2} \|\Psi_k\| < \infty \quad \textit{and} \quad c := \sup_{\omega \in [0, \pi[} \|\mathbf{F}_\omega^{-1}\| < \infty, \tag{3.20}$$

where $\mathbf{F}_\omega$, denotes the spectral density matrix of $(\boldsymbol{\varepsilon}_t)_{t \in \mathbb{Z}}$ (see hereafter). We will further assume that for some estimator $\widehat{\mathbf{F}}_\omega$ we have

$$\max_{1 \leq j \leq q} \|\widehat{\mathbf{F}}_{\omega_j} - \mathbf{F}_{\omega_j}\| = o_P(n^{-1/4}). \tag{3.21}$$

We note that (3.21) can be obtained under mild assumptions. See e.g. Theorem 2.1 in Politis (2011). We recall that the spectral density matrix of a multivariate process $(\mathbf{X}_t)_{t \in \mathbb{Z}}$ is defined as

$$\mathbf{F}_\omega = \sum_{h \in \mathbb{Z}} \mathbf{C}_h e^{-\mathrm{i}t\omega} \qquad \forall\, \omega \in [-\pi, \pi],$$

where $\mathbf{C}_h = \mathrm{Cov}(\mathbf{X}_h, \mathbf{X}_0)$. A simple estimator of $\mathbf{F}_\omega$, is given by

$$\widehat{\mathbf{F}}_{\omega_j} = \sum_{|h| < \ell_n} \left(1 - \frac{|h|}{\ell_n}\right) \mathbf{C}_h e^{-it\omega_j} \quad 1 \leq j \leq q,$$

for some appropriate sequence $\ell_n \nearrow \infty$. We tacitly assume in the following that ℓ_n is such that the corresponding estimator is consistent, see e.g. Brillinger (1969) or Politis (2011) for more details. Then we define the test statistic

$$L_n^{\mathbf{X}} := \max_{1 \leq j \leq q} \|\widehat{\mathbf{F}}_{\omega_j}^{-1/2} \mathcal{X}_n^{\mathbf{X}}(\omega_j)\|^2 = \max_{1 \leq j \leq q} \mathrm{tr}\Big(\widehat{\mathbf{F}}_{\omega_j}^{-1} I_n^{\mathbf{X}}(\omega_j)\Big),$$

where A^* is the conjugate transpose of a matrix A. In order to get a test that is valid under Assumption 6, we need the following result.

Theorem 11. *Let $c_n^p = \log(n/2) + (p-1)\log\log(n/2) - \log[(p-1)!]$. Then under Assumption 6 and under $\mathcal{H}_0$ we have that $L_n^{\mathbf{X}} - c_n^p \xrightarrow{d} \mathcal{G}$.*

We note that in this theorem we do not require distinct eigenvalues of C. The following result is now immediate:

Test 2. *Under Assumption 6, the test that rejects $\mathcal{H}_0$ if*

$$L_n^{\mathbf{X}} - c_n^p > Q_{1-\alpha}^{\mathcal{G}}, \tag{3.22}$$

has asymptotic level α.

Note that this test is not a generalization of Test 1. This new test attributes the same importance to each eigenvector $v_1, \dots, v_p$ of the covariance matrix $\mathbf{C}_0 = \mathrm{Var}(\mathbf{X})$. Indeed, even under independence $\mathbf{F}_\omega = \mathbf{C}_0$, we have that

$$\|\widehat{\mathbf{F}}_{\omega_j}^{-1/2}\mathcal{X}_n^{\mathbf{X}}(\omega_j)\|^2 = \left\| n^{-1/2}\sum_{t=1}^{n} \widehat{\mathbf{C}}_0^{-1/2}\mathbf{X}_t e^{-\mathrm{i}t\omega}\right\|^2 \neq \|\mathcal{X}_n^{\mathbf{X}}(\omega)\|^2,$$

so in general, $L_n^{\mathbf{X}} \neq M_n^{\mathbf{X}}$.

3.3.2 Functional test

Now we would like to take into account the infinite dimensional character of functional data. Here we limit ourselves again to Assumption 5. Then the following results are immediate applications of Theorems 9 and 10.

Test 3. *Under Assumption 5, the test that rejects $\mathcal{H}_0$ if*

$$\hat{\lambda}_1^{-1}(M_n^{X,p_n} - \hat{b}_n^{p_n}) > Q_{1-\alpha}^{\mathcal{G}}, \tag{3.23}$$

has asymptotic level α, for any sequence $(p_n)_{n\geq 1}$ that satisfies conditions (3.8).

Test 4. *Under Assumption 5 and the assumptions of Theorem 10, the test that rejects $\mathcal{H}_0$ if*

$$\hat{\lambda}_1^{-1}(M_n^{X} - \hat{b}_n) > Q_{1-\alpha}^{\mathcal{G}}, \tag{3.24}$$

has asymptotic level α.

Remark 3. Note that we are interested in detecting large values of $\|\mathcal{X}(\omega_j)\|^2$. This is why we only consider the right tail of the Gumbel distribution in our tests.

3.4 Proofs of Chapter 3

3.4.1 Proofs of the main results

Proof of Lemma 9. Note that

$$b_n^p = \lambda_1 \log(n/2) \; - \; \lambda_1 \sum_{j=2}^{p} \log(1-\lambda_j/\lambda_1)$$

and $\log(1-\lambda_j/\lambda_1) \leq \lambda_j/\lambda_1$. The result follows from $\sum_{j\geq 1}\lambda_j = \|C\| < \infty$. ☐

Proof of Lemma 10. It is well known that

$$\sup_{j\geq 1}|\lambda_j-\widehat{\lambda}_j|\leq \|C-\widehat{C}\|=O_P(n^{-1/2}). \tag{3.25}$$

In particular, (3.25) yiels that $|1/\lambda_1-1/\widehat{\lambda}_1|\xrightarrow{P}0$. By elementary algebra we get

$$\frac{M_n^p-b_n^p}{\lambda_1}-\frac{M_n^p-\widehat{b}_n^p}{\widehat{\lambda}_1}=(M_n^p-b_n^p)\left(\frac{1}{\lambda_1}-\frac{1}{\widehat{\lambda}_1}\right)+\frac{\widehat{b}_n^p-b_n^p}{\widehat{\lambda}_1}.$$

The weak convergence of $M_n^p-b_n^p$ thus implies that the first term tends to zero in probability. Next,

$$\widehat{b}_n^p-b_n^p=(\widehat{\lambda}_1-\lambda_1)\log(n/2)\;-\;\sum_{j=2}^{p}\Big(\widehat{\lambda}_1\log(1-\widehat{\lambda}_j/\widehat{\lambda}_1)-\lambda_1\log(1-\lambda_j/\lambda_1)\Big) \tag{3.26}$$

From (3.25) it is immediate that $(\widehat{\lambda}_1-\lambda_1)\log(n/2)\to 0$. For the summands in the second term we apply the mean-value theorem term by term and use (3.25), $\lambda_1/\lambda_j>1$ when $j\geq 1$, and $p=o(n^{1/2})$. □

Proof of Theorem 8. We assume that $E\|X_1\|^s<\infty$, for some $s>2$. Here $p\geq 1$ is arbitrary but fixed. Let $(Z_t)_{t\in\mathbb{Z}}$ be an H-valued sequence of Gaussian variables with $\mathrm{Var}(Z_1)=\mathrm{Var}(X_1)$ and denote its DFT by $\widetilde{\mathcal{X}}_n(\omega)$, its periodogram by $\widetilde{I}_n(\omega)$, etc. By basic properties of the Gaussian law we get

$$\left\{\|\widetilde{\mathcal{X}}_n^p(\omega_j)\|^2,\ j=1,\dots,q\right\}\overset{d}{=}\left\{\sum_{k=1}^{p}\lambda_k E_{kj},\ j=1,\dots,q\right\},$$

where the E_{kj} are i.i.d. standard exponential variables. The variables $Z_j^p:=\sum_{k=1}^p\lambda_kE_{kj}$ then follow a hypo-exponential distribution with parameters $\lambda_1,\dots,\lambda_p$. Since we impose $\lambda_k>\lambda_{k+1}$ for any $k\geq 1$, we can get (see e.g. Kang and Serfolo (1999)) an explicit expression for the distribution function

$$F^{(p)}(x)=P(Z_j^p\leq x)=1-\sum_{k=1}^{p}\alpha_{k,p}e^{-x/\lambda_k},\quad\text{where}\quad \alpha_{k,p}=\prod_{\substack{j=1\\ j\neq k}}^{p}\frac{1}{1-\lambda_j/\lambda_k}.$$

We deduce that

$$\begin{aligned}
P\left(\lambda_1^{-1}(\widetilde{M}_n^p-b_n^p)\leq x\right)&=F^{(p)}(\lambda_1x+b_n^p)^q\\
&=\left(1-\sum_{k=1}^{p}\alpha_{k,p}e^{-\frac{\lambda_1}{\lambda_k}x}\left(\frac{2}{n\alpha_{1,p}}\right)^{\frac{\lambda_1}{\lambda_k}}\right)^q\\
&=\left(1-\frac{2}{n}\left\{\sum_{k=2}^{p}\alpha_{k,p}\left(\frac{e^{-x}}{\alpha_{1,p}}\right)^{\lambda_1/\lambda_k}\left(\frac{2}{n}\right)^{\lambda_1/\lambda_k-1}+e^{-x}\right\}\right)^q. &(3.27)
\end{aligned}$$

Clearly,

$$\sum_{k=2}^{p}\alpha_{k,p}\left(\frac{e^{-x}}{\alpha_{1,p}}\right)^{\lambda_1/\lambda_k}\left(\frac{2}{n}\right)^{\lambda_1/\lambda_k-1}\to 0. \tag{3.28}$$

Since $q \sim n/2$ we can obtain indeed that (3.27) converge to the Gumbel distribution function.

Next we want to transfer the result to our original data. Clearly, we have that

$$\left|P\Big((M_n^p - b_n^p)/\lambda_1 \le x\Big) - e^{-e^{-x}}\right| \le \rho_n^p + \left|P\Big((\widetilde{M}_n^p - b_n^p)/\lambda_1 \le x\Big) - e^{-e^{-x}}\right|, \tag{3.29}$$

where

$$\rho_n^p := \sup_{x\in\mathbb{R}} |P(M_n^p \le x) - P(\widetilde{M}_n^p \le x)|. \tag{3.30}$$

Under the stronger assumption that $E\|X_1\|^4 < \infty$, we can deduce from Proposition 2 that $\rho_n^p \to 0$ when $n \to \infty$ for any fixed $p \ge 1$.

In order to prove the result when $E\|X_1\|^s < \infty$, for some $s > 2$, we use a truncation argument that is detailed in Section 3.4.5. The proof then followed from Corollary 2. □

Proof of Theorem 9. We may proceed as in the proof of Theorem 8. We need to verify (3.28) with and (3.30) with p_n instead of fixed p. As of (3.28) we refer to Lemma 13, while for (3.30) we may still employ Proposition 2. □

Proof of Theorem 10. We start by noting that

$$\lambda_1^{-1}(M_n - b_n) = \lambda_1^{-1}(M_n - M_n^{p_n}) + \lambda_1^{-1}(M_n^{p_n} - b_n^{p_n}) + \lambda_1^{-1}(b_n^{p_n} - b_n).$$

The third term converges to zero by Lemma 9, and the second term converges weakly to a Gumbel random variable under the assumptions of Theorem 9. Hence, by Slutzky's lemma, weak convergence of $\lambda_1^{-1}(M_n - b_n)$ to a Gumbel distribution holds if we can verify that the first term tends to zero in probability.

To this end, we define

$$\delta_j = \|\mathcal{X}_n(\omega_j)\|^2 - \|\mathcal{X}_n^p(\omega_j)\|^2 = \sum_{k>p} \frac{1}{n}\left|\sum_{t=1}^n \langle X_t, v_k\rangle e^{-\mathrm{i}t\omega_j}\right|^2.$$

Then for any $a > 0$ we have

$$\begin{aligned}
P(|M_n - M_n^p| > a) &= P(M_n - M_n^p > a) \\
&= P\left(\max_{j=1,\dots,q}\left\{\|\mathcal{X}^p(\omega_j)\|^2 + \delta_j\right\} - M_n^p > a\right) \\
&\le P\left(\max_{j=1,\dots,q} \delta_j > a\right) \\
&\le \sum_{j=1}^q \sum_{k>p} P\left(\frac{1}{n}\left|\sum_{t=1}^n \langle X_t, v_k\rangle \cos(t\omega_j)\right|^2 > a\ell_k/2\right) \\
&\quad + \sum_{j=1}^q \sum_{k>p} P\left(\frac{1}{n}\left|\sum_{t=1}^n \langle X_t, v_k\rangle \sin(t\omega_j)\right|^2 > a\ell_k/2\right).
\end{aligned}$$

Next we recall that if ξ_t are independent random variables with $E|\xi_t|^s$ for some $s > 2$ and if $E\xi_t = 0$, then by Rosenthal's inequality (see e.g. Rosenthal (1970)) there is a constant C_s depending only on s such that

$$E\left|\sum_{t=1}^n \xi_t\right|^s \le C_s\left[\sum_{t=1}^n E|\xi_t|^s + \left(\sum_{t=1}^n E|\xi_t|^2\right)^{s/2}\right].$$

Applying this inequality to $\xi_t = \langle X_t, v_k \rangle \cos(t\omega_j)$ in combination with the Markov inequality, we deduce that

$$\begin{aligned}
&P\left(\frac{1}{n}\left|\sum_{t=1}^{n}\langle X_t, v_k\rangle \cos(t\omega_j)\right|^2 > a\ell_k/2\right) \\
&\quad\le C_s\,(na\ell_k/2)^{-s/2}\left[\sum_{t=1}^{n} E\,|\langle X_t, v_k\rangle|^s + \left(\sum_{t=1}^{n} E\,|\langle X_t, v_k\rangle|^2\right)^{s/2}\right] \\
&\quad\le C_s\,(na\ell_k/2)^{-s/2}\left[nE\,|\langle X_1, v_k\rangle|^s + (n\lambda_k)^{s/2}\right] \\
&\quad\le C_s(2/a)^{s/2}\left[n^{1-s/2}\ell_k^{-s/2}E\,|\langle X_1, v_k\rangle|^s + (\lambda_k/\ell_k)^{s/2}\right].
\end{aligned}$$

When $\xi_t = \langle X_t, v_k \rangle \sin(t\omega_j)$ we use the same argument. Summation over j and k yields the result due to the conditions (3.11) and (3.12). $\square$

Below we write $a_n \ll b_n$ to indicate that $\limsup_n |a_n/b_n| < \infty$.

Proof of Corollary 1. First remark that applying twice the Cauchy–Schwarz inequality, we get that

$$E\,|\langle X_1, v_k\rangle|^s \le \lambda_k^{1/2}\left(E\|X_1\|^{2(s-1)}\right)^{1/2}.$$

Thus, a sufficient condition for (3.12) to hold, is that

$$\sum_{k>p_n} \ell_k^{-s/2}\lambda_k^{1/2} = o(n^{s/2-1}) \tag{3.31}$$

We first consider setting (**E1**), i.e. $\lambda_k \sim \rho^k$ with $0<\rho<1$. We choose ℓ_k proportional to k^{-2}. We show that we may choose $p_n = \lfloor c\log(n)\rfloor$. Indeed, for any $\rho < \rho_0 < 1$, we have that

$$\sum_{k>p_n} \ell_k^{-s/2}\lambda_k^{1/2} \ll \sum_{k>p_n} k^{s/2}\rho^{k/2} \ll \sum_{k>p_n} \rho_0^{k/2} \ll \rho_0^{p_n/2} \ll n^{-\frac{c}{2}\log\left(1/\rho_0\right)} = o(1),$$

so (3.31) is clearly satisfied without further restrictions. Similarly, we find that

$$\sum_{k>p_n} (\lambda_k/\ell_k)^{s/2} \ll \rho_0^{sp_n/2} \ll n^{-\frac{sc}{2}\log\left(1/\rho_0\right)} = o(1/n),$$

whenever $c > 2/(s\log(1/\rho))$ and ρ_0 is close enough to ρ.

Next we turn to condition (3.8). While the second part is obvious, we need for the first part that

$$\frac{p_n^3}{\lambda_{p_n}^{1/2}} \ll \log^3(n)n^{\frac{c}{2}\log(1/\rho)} = o\left(\frac{n^{1/6}}{\log^{7/6}(n)}\right),$$

which follows if $c < 1/(3\log(1/\rho))$. Thus, we need to choose c such that

$$\frac{2}{s\log(1/\rho)} < c < \frac{1}{3\log(1/\rho)}.$$

This is possible whenever $s > 6$.

We now turn to the case (**E2**), i.e. $\lambda_k \sim k^{-\nu}$ with $\nu > 1$. We choose ℓ_k proportional to $k^{-1}\log^{-2}(k)$ and show that we may set $p_n = \lfloor n^\beta \rfloor$, for some appropriate $\beta > 0$. Indeed, for any $1 < \tilde{\nu} < \nu$, we have that

$$\sum_{k>p_n} \lambda_k^{1/2}/\ell_k^{s/2} \ll \sum_{k>p_n} k^{\frac{s}{2}-\frac{\nu}{2}} \log^s(k) \ll \sum_{k>p_n} k^{\frac{s}{2}-\frac{\tilde{\nu}}{2}} \ll n^{\beta\left(\frac{s}{2}-\frac{\tilde{\nu}}{2}+1\right)}.$$

so (3.31) is satisfied whenever $\beta < \frac{s-2}{2+s-\nu}$ if $s > \nu+2$ or without further constraints if $s \leq \nu + 2$. Similarly, we have that

$$\sum_{k>p_n} (\lambda_k/\ell_k)^{s/2} \ll \sum_{k>p_n} k^{\frac{s}{2}-\frac{s\tilde{\nu}}{2}} \ll n^{\beta\left(\frac{s}{2}-\frac{s\tilde{\nu}}{2}+1\right)}.$$

This term is $o(1/n)$ if $\tilde{\nu}$ is chosen close enough to ν, $s > \frac{2}{\nu-1}$ and if $\beta > \frac{1}{(\nu-1)s/2-1}$.

Then we turn to condition (3.8). The first part is satisfied if

$$\frac{p_n^3}{\lambda_{p_n}^{1/2}} \ll n^{\beta(3-\nu/2)} = o\left(\frac{n^{1/6}}{\log^{7/6}(n)}\right).$$

It requires $\beta < \frac{1}{6\,(\nu/2+3)}$. For the second part we require $\beta < \min\left\{\min_{k\geq 2} \frac{1}{k}\left(\frac{\lambda_1}{\lambda_k}-1\right), 1\right\}$. Thus, β must be such that

$$\frac{1}{(\nu-1)s/2-1} < \beta < \min\left\{\frac{s-2}{2+s-\nu}, \frac{1}{6\,(\nu/2+3)}, \min_{k\geq 2}\frac{1}{k}\left(\frac{\lambda_1}{\lambda_k}-1\right), 1\right\}.$$

□

3.4.2 Normal approximation

A key step in our proof is to find a bound for ρ_n^p, as defined in (3.30). This quantity determines the distance between the distribution function of the statistic of interest computed from our sample, and from the same statistic computed from a normal random sample. To this end, we will apply a powerful result of Chernozhukov et al. (2017). They provide a Gaussian approximation of $P(S_n^{\mathbf{X}} \in A)$, where $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are i.i.d. random vectors, $S_n^{\mathbf{X}} = n^{-1/2}\sum_{i=1}^n \mathbf{X}_i$, and A runs through the class $\mathcal{A}^{\mathrm{sp}}(s)$ of s–sparsely convex subsets of $\mathbb{R}^N$, i.e. finite intersections of sets whose indicator function depends on at most s–components. More precisely, they provide a bound for

$$\rho_n(\mathcal{A}^{\mathrm{sp}}(s)) = \sup_{A\in\mathcal{A}^{\mathrm{sp}}(s)} |P(S_n^{\mathbf{X}} \in A) - P(S_n^{\mathbf{Y}} \in A)|,$$

where $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ denotes a N variate i.i.d. Gaussian sequence with same variance as $\mathbf{X}$. They make the following assumptions (we use their numbering)

$$\text{(M.1”)} \quad \frac{1}{n}\sum_{t=1}^n E|u'\mathbf{X}_t|^2 \geq b \quad \text{for all} \;\; u \in \mathbb{S}^{N-1} \;\; \text{and} \;\; \|u\|_0 \leq s;$$

$$\text{(M.2)} \quad \frac{1}{n}\sum_{t=1}^n E|\mathbf{X}_{t,j}|^{2+k} \leq B_n^k \quad \text{for all} \;\; j = 1, \ldots, d, \quad k = 1, 2;$$

$$\text{(E.1)} \quad E\exp(|\mathbf{X}_{t,j}|/B_n) \leq 2, \quad \text{for all} \;\; t = 1, \ldots, n \quad \text{and} \quad j = 1, \ldots, N;$$

$$\text{(E.2)} \quad E\max_{1\leq j\leq N}(|\mathbf{X}_{t,j}|/B_n)^4 \leq 2, \quad \text{for all} \quad t = 1, \ldots, n,$$

where $B_n \geq 1$ is a deterministic sequence, possibly converging to infinity. Under conditions (M.1") and (M.2), they show in their Proposition 3.2 that

$$\rho_n(\mathcal{A}^{\mathrm{sp}}(s)) \leq C \cdot \frac{B_n^{1/3} \log^{7/6}(N \cdot n)}{n^{1/6}}, \quad \text{or} \quad \rho_n(\mathcal{A}^{\mathrm{sp}}(s)) \leq C' \cdot \frac{B_n^{2/3} \log^{7/6}(N \cdot n)}{n^{1/6}}, \tag{3.32}$$

whether condition (E.1) or (E.2) is satisfied. The constants C and C', only depends on b and s. This dependence is crucial for our application, however it is not explicitly stated in Chernozhukov et al. (2017). We thus provide it in equation (3.47) at the end of Section 3.4.3.

In the next proposition, we show how these results can be applied to our setting. At this point, we only use the second bound from (3.32). Note that the first bound necessitates moments of all order through condition (E.1). On the other hand, the dependence on B_n is slightly milder, which will be useful in Section 3.4.5 where we will consider a truncated process.

Proposition 2. *Let ρ_n^p be given as in* (3.30). *We suppose that $E\|X_1\|^4 < \infty$, and set $B = E\|X_1\|^3 \vee (E\|X_1\|^4)^{1/2}$. Then there exists a universal constant K such that*

$$\rho_n^p \;\leq\; K \cdot B^{2/3} \, \frac{p^3 \, \log^{7/6}(pn^2)}{\lambda_p^{1/2} \, n^{1/6}}. \tag{3.33}$$

Proof of Proposition 2. We define

$$\begin{aligned} \xi_t^{(p)} &= (\langle X_t, v_1 \rangle, \dots, \langle X_t, v_p \rangle)', \\ W_t^{(q)} &= (\cos(t\omega_1), \sin(t\omega_1), \dots, \cos(t\omega_q), \sin(t\omega_1))' \end{aligned}$$

and

$$\xi_t^{(p)} \otimes W_t^{(q)} = (\langle X_t, v_1 \rangle W_t^{(q)}, \dots, \langle X_t, v_p \rangle W_t^{(q)})'.$$

The vectors $\mathbf{X}_t := \xi_t^{(p)} \otimes W_t^{(q)}$ are independent random vectors in $\mathbb{R}^{2pq}$. Next we define the partial sums $S_n^{\xi \otimes W} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{X}_t$ and the sets $J_k = J + 2k$, where

$$J = \{1, 2, 2q+1, 2q+2, \dots, 2(p-1)q+1, 2(p-1)q+2\}$$

and $k = 0, \dots, q-1$. We note that

$$\begin{aligned} P(M_n^p \leq x) &= P\left(\|\mathcal{X}_n^p(\omega_j)\|^2 \leq x, \, j = 1, \dots, q \right) \\ &= P(\|(S_n^{\xi \otimes W})_{j \in J_k}\|^2 \leq x, \, 0 \leq k \leq q-1). \end{aligned}$$

The sets of the form

$$A = \{\omega \in \mathbb{R}^{2pq} \colon \sum_{j \in J_k} \omega_j^2 \leq x, 0 \leq k \leq q-1\}$$

are $2p$-sparsely convex sets. We can thus apply (3.47) but where p is replaced by $2pq$ and s by $2p$. Indeed, since $\|\mathbf{X}_{t,j}\| \leq \|X_t\|$ for all $j = 1, \dots, 2pq$, it is easy to see that (E.2) and (M.2) are satisfied with $B_n = B$ and condition (M.1") follows from Lemma 11 hereafter. □

Lemma 11. *We define* $\mathbf{X}_t := \xi_t^{(p)} \otimes W_t^{(2q)}$. *We have that*

$$\frac{\lambda_p}{2} \le \frac{1}{n}\sum_{t=1}^{n} E|u'\mathbf{X}_t|^2 \le \frac{\lambda_1}{2} \tag{3.34}$$

for all $u \in \mathbb{S}^{2pq-1}$, *and in particular when* $\|u\|_0 \le 2p$.

Proof of Lemma 11. We denote $u = (\mathbf{u}_1', \ldots, \mathbf{u}_p')' \in \mathbb{R}^{2pq}$ with $\mathbf{u}_k' \in \mathbb{R}^{2q}$. We have that

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} E|u'\mathbf{X}_t|^2 &= \frac{1}{n}\sum_{t=1}^{n}\sum_{j,k=1}^{p} E\langle X_t, v_j\rangle\langle X_t, v_k\rangle \boldsymbol{u}_j' W_t^{(q)} \boldsymbol{u}_k' W_t^{(q)} \\ &= \sum_{j=1}^{p} \lambda_j \mathbf{u}_j' \frac{1}{n}\sum_{t=1}^{n} W_t^{(q)} (W_t^{(q)})' \mathbf{u}_j.\end{aligned}$$

Each coefficient of the matrix $\mathbf{W} = \frac{1}{n}\sum_{t=1}^{n} W_t^{(q)} \otimes W_t^{(q)}$ have one of the following forms

$$\frac{1}{n}\sum_{t=1}^{n}\sin(t\omega_k)\cos(t\omega_\ell) = \frac{1}{2n}\sum_{t=1}^{n}\big(\sin(t\omega_k + t\omega_\ell) + \sin(t\omega_k - t\omega_\ell)\big), \tag{3.35}$$

$$\frac{1}{n}\sum_{t=1}^{n}\sin(t\omega_k)\sin(t\omega_\ell) = \frac{1}{2n}\sum_{t=1}^{n}\big(\cos(t\omega_k - t\omega_\ell) - \cos(t\omega_k + t\omega_\ell)\big), \tag{3.36}$$

$$\frac{1}{n}\sum_{t=1}^{n}\cos(t\omega_k)\cos(t\omega_\ell) = \frac{1}{2n}\sum_{t=1}^{n}\big(\cos(t\omega_k - t\omega_\ell) + \cos(t\omega_k + t\omega_\ell)\big), \tag{3.37}$$

for some $1 \le k, \ell \le \lfloor (n-1)/2 \rfloor$. We see that (3.35) is always zero whereas both (3.36) and (3.37) are equal to $1/2$ if $\omega_k \pm \omega_\ell \in 2\pi\mathbb{Z}$ (i.e. when $k = \ell$) and to zero if not. In other words, $\mathbf{W}$ is just the identity matrix times $1/2$. Thus,

$$\frac{1}{n}\sum_{t=1}^{n} E|u'\mathbf{X}_t|^2 = \frac{1}{2}\sum_{j=1}^{p}\lambda_j \|\mathbf{u}_j\|^2. \tag{3.38}$$

The sum is maximized if $\|\mathbf{u}_1\| = 1$ and minimized if $\|\mathbf{u}_p\| = 1$. □

3.4.3 Explicit constants in the highdimensional CLT

The constants C and C' in (3.32) depends on the parameters b and s (in our setting, this correspond to λ_p and $2p$). If we want to let p grow to infinity, we need to make this dependence explicit. This is the purpose of this section. The next Lemma is a reformulation of Lemma A.1 in Chernozhukov et al. (2017) where we have made more explicit the dependence on the variance lower bound. This anti-concentration inequality is the key ingredient in Chernozhukov et al. (2017) and it is essentially due to Nazarov (2003). For completeness we provide a detailed proof of it. The main arguments have been kindly communicated to us by Kengo Kato.

Lemma 12. *Let* $Y \sim \mathcal{N}_p(0, \Sigma)$ *be such that* $EY_j^2 \ge b$, *for all* $j = 1, \ldots, p$, *for some constant* $b > 0$. *Then for every* $y \in \mathbb{R}^p$ *and* $\delta > 0$, *we have that*

$$P(Y \le y + \delta) - P(Y \le y) \le \frac{\delta}{b^{1/2}}\left(\sqrt{2\log p} + 2\right). \tag{3.39}$$

The inequalities between vectors are coordinatewise.

Proof of Lemma 12. Firstly, by considering $X = \Sigma^{-1/2}Y$, we remark that it is equivalent to prove that for all $X \sim \mathcal{N}_p(0, I_p)$, we have that

$$P(X \in K(\delta) \setminus K) \leq \delta \left(\sqrt{2\log p} + 2\right), \tag{3.40}$$

where

$$K(\delta) = \{x \in \mathbb{R}^p,\ a_j' x \leq b_j + \delta,\ \forall j = 1, \dots, p\}, \quad K = K(0). \tag{3.41}$$

and $a_j \in \mathbb{S}^{p-1}$ and $b_j \in \mathbb{R}$, for $j = 1, \dots, p$. We now define $G(t) = P(X \in K(t))$, which is in fact the cdf of the random variable $\max_{1 \leq j \leq p}\{a_j' X - b_j\}$. Thus, we know that G is absolutely continuous and hence is almost everywhere differentiable with

$$P(X \in K(t) \setminus K) = G(\delta) - G(0) = \int_0^\delta G_+'(t)dt, \tag{3.42}$$

where G_+' denotes the right derivative of G. Actually we will show that it is *everywhere* right differentiable and that $G_+'(t) \leq \sqrt{2\log p} + 2$, for all $t \in \mathbb{R}$, which according to (3.42), would complete the proof the Lemma. Note that by replacing b_j by $b_j + t$, we see that it is enough to prove that G is right differentiable at $t = 0$ and that $G_+'(0) \leq \sqrt{2\log p} + 2$, i.e. that

$$\Gamma_p(K) \ = \ \lim_{\delta \searrow 0} \frac{P\left(X \in K(\delta) \setminus K\right)}{\delta} \ \leq \ \sqrt{2\log p} + 2. \tag{3.43}$$

Note that Γ_p is the standard Gaussian surface measure. In order to prove (3.43), we introduce some additional notations. For $x \in \mathbb{R}^p$, $P_K x$ denotes the projection onto the convex polyhedron K. For a face F of K, we define

$$N_F = \{x \in \mathbb{R}^p \setminus K,\ P_K x \in \operatorname{relint}(F)\}, \quad N_F(\delta) = N_F \cap K(\delta),$$

where $\operatorname{relint}(F)$ denotes the relative interior of F. Note that $K(\delta) \setminus K = \bigcup_{F:\text{face of } K} N_F(\delta)$. Now, for any face F of K with dimension at most $p-2$, we have that $N_F(\delta) \subset \{x \in \mathbb{R}^p,\ \operatorname{dist}(x, F) \leq C\delta\}$, where $\operatorname{dist}(x, F)$ is the Euclidean distance between x and F, and C is a constant that is proportional to $\sqrt{p}$. This shows that when computing (3.43), we can restrict to the facet of K (i.e. face of dimension $p-1$). Indeed, we have that

$$P\left(X \in K(\delta) \setminus K\right) = P\Big(X \in \bigcup_{F:\text{facet of } K} N_F(\delta)\Big) + O(\delta^2).$$

Then, by similar arguments as in Nazarov (2003) we can show that for any facet F of K,

$$\begin{aligned} \Gamma_p(K) &= \lim_{\delta \searrow 0} \frac{P\left(X \in N_F(\delta)\right)}{\delta} \\ &= \int_F \varphi_p(x) d\sigma(x) \ \leq \ (\operatorname{dist}(0, F) + 1) \cdot P\left(X \in N_F\right), \end{aligned} \tag{3.44}$$

where $\varphi_p(x) = (2\pi)^{-p/2} e^{-\|x\|^2/2}$ and $d\sigma$ denotes the standard surface measure. Intuitively (3.44) means that the points $x \in \partial K$ whose tangent to K is close to the origin do not contribute much to the Gaussian surface of K. Finally we have that

$$\begin{aligned}
\lim_{\delta\searrow 0} \frac{P\left(X\in K(\delta)\setminus K\right)}{\delta} &= \sum_{F:\,\text{facet of }K} \int_F \varphi_p(x)d\sigma(x)\\
&= \sum_{\operatorname{dist}(0,F)\le\sqrt{2\log(p)}} \int_F \varphi_p(x)d\sigma(x) + \sum_{\operatorname{dist}(0,F)>\sqrt{2\log(p)}} \int_F \varphi_p(x)d\sigma(x)\\
&\le \left(\sqrt{2\log p}+1\right) \sum_{\operatorname{dist}(0,F)\le\sqrt{2\log(p)}} P\left(X\in N_F\right) + \sum_{\operatorname{dist}(0,F)>\sqrt{2\log(p)}} \int_{\mathbb{R}^{p-1}} \varphi_{p-1}(y)d\sigma(y).
\end{aligned}$$

In the last inequality, we used (3.44) for the first term. Whereas for the second term, we used the fact that all $x\in F$ can be decompose as $x=\operatorname{dist}(0,F)v+x_0$, where v is a unit normal vector to ∂K at F and $x_0\in v^\perp\cong\mathbb{R}^{p-1}$. We then have that $\|x\|^2=\operatorname{dist}(0,F)^2+\|x_0\|^2>2\log(p)+\|x_0\|^2$, which yields

$$\varphi_p(x)=(2\pi)^{-p/2}e^{-\|x\|^2/2}<p^{-1}(2\pi)^{-1/2}\cdot(2\pi)^{-(p-1)/2}e^{-\|x_0\|^2/2}=p^{-1}(2\pi)^{-1/2}\cdot\varphi_{p-1}(y),$$

for some $y\in\mathbb{R}^{p-1}$. To conclude, recall that from (3.41) we know that K has at most p facets, and $(2\pi)^{-1/2}\le 1$. □

In the sequel, we give explicit dependence in b of the bounds obtained in **Lemma 5.1**, **Corollary 5.1**, **Theorem 2.1**, **Corollary 5.1** and **Proposition 2.1** in Chernozhukov et al. (2017). We will basically add a factor $b^{-1/2}$ in front of every terms that comes out of Nazarov's inequality i.e. Lemma 12 above. We will denote by K_i for $i\ge 1$, a genuine universal constant i.e. independent of n, p and b. On the other hand the variables B_n, $\mathcal{A}^{\text{re}}$, L_n, $\overline{L}_n$, $M_{n,X}(\phi)$, $M_n(\phi)$ and $\widetilde{\mathbf{X}}$ are defined as in Chernozhukov et al. (2017). Furthermore in this section, $\mathbf{X}_1,\dots,\mathbf{X}_n$ denotes a p variate sequence of independent random vectors, that will actually correspond in our setting to $\xi_t^{(p)}\otimes W_t^{(2q)}$ for $t=1,\dots,n$. Then $\mathbf{Y}_1,\dots,\mathbf{Y}_n$ denotes a p variate i.i.d. Gaussian sequence with same variance as $\mathbf{X}$. The authors further assume condition (M.2) and

$$\text{(M.1)}\qquad \frac{1}{n}\sum_{t=1}^n E|\mathbf{X}_{t,j}|^2\ge b\quad\text{for all}\;\; j=1,\dots,p.$$

A meticullous reading of **Lemma 5.1**'s proof in Chernozhukov et al. (2017) yields

$$\begin{aligned}
\varrho_n &= \sup_{y\in\mathbb{R},v\in[0,1]} |P(\sqrt{v}S_n^{\mathbf{X}}+\sqrt{1-v}S_n^{\mathbf{Y}}\le y)-P(S_n^{\mathbf{Y}}\le y)|\\
&\le K_1\left\{\frac{\phi^2\log^2 p}{n^{1/2}}\left(\phi L_n\varrho_n+L_n\frac{\log^{1/2}p}{b^{1/2}}+\phi M_n(\phi)\right)+\frac{\log^{1/2}p}{\phi\, b^{1/2}}\right\}.
\end{aligned}$$

Then **Corollary 5.1** provides the same result for arbitrary hyperrectangles except that p must be replaced by $2p$ in the bound and by noting that $M_{n,\widetilde{\mathbf{X}}}(\phi)\le M_{n,\mathbf{X}}(2\phi)$ whenever $p\ge 2$. To summarize we have that

$$\begin{aligned}
\varrho_n' &= \sup_{A\in\mathcal{A}^{\text{re}},v\in[0,1]} |P(\sqrt{v}S_n^{\mathbf{X}}+\sqrt{1-v}S_n^{\mathbf{Y}}\in A)-P(S_n^{\mathbf{Y}}\in A)|\\
&\le K_2\left\{\frac{\phi^3\log^2 p}{n^{1/2}}L_n\varrho_n'+\frac{\phi^2\log^{3/2}(p)}{b^{1/2}n^{1/2}}L_n+\frac{\phi^3\log^2 p}{n^{1/2}}M_n(2\phi)+\frac{\log^{1/2}p}{b^{1/2}\phi}\right\}.
\end{aligned}$$

For **Theorem 2.1** we apply **Corollary 5.1** with $\phi = \phi_n/2 = \frac{1}{2K_2 \vee 1} \frac{n^{1/6}}{\overline{L}_n^{1/3} \log^{2/3} p}$ and get that

$$\varrho_n' \le K_2 \Bigg\{ \frac{1}{8(K_2 \vee 1)^3} \frac{L_n}{\overline{L}_n} \varrho_n' + \frac{1}{4(K_2 \vee 1)^2} \frac{L_n}{\overline{L}_n^{2/3}} \frac{\log^{1/6}(p)}{b^{1/2} n^{1/6}} + \frac{1}{8(K_2 \vee 1)^3} \frac{M_n(\phi_n)}{\overline{L}_n} + 2(K_2 \vee 1) \frac{\log^{7/6} p \, \overline{L}_n^{1/3}}{b^{1/2} n^{1/6}} \Bigg\}.$$

Therefore we get that

$$\rho_n(\mathcal{A}^{\mathrm{re}}) \le K_3 \left\{ \frac{M_n(\phi_n)}{\overline{L}_n} + \frac{\log^{7/6} p \, \overline{L}_n^{1/3}}{b^{1/2} \, n^{1/6}} \right\}.$$

For **Proposition 2.1** whether condition (E.1) or (E.2) is satisfied, we obtain that

$$\rho_n(\mathcal{A}^{\mathrm{re}}) \le K_4 \, \frac{B_n^{1/3} \, \log^{7/6}(pn)}{b^{1/2} \, n^{1/6}} \quad \text{or} \quad \rho_n(\mathcal{A}^{\mathrm{re}}) \le K_5 \, \frac{B_n^{2/3} \, \log(pn)}{b^{1/2} \, n^{1/6}}. \tag{3.45}$$

Let $a, d > 0$ be some constants. The authors defined $\mathcal{A}^{\mathrm{si}}(a,d)$ the class of Borel sets A in $\mathbb{R}^p$ such that there exists an m-generated convex set A^m (i.e. an m convex polytope) such that $A^m \subset A \subset A^{m,\varepsilon}$, where $\varepsilon = a/n$. (By m-polytope we mean that A^m can be written as $\bigcap_{v \in \mathcal{V}(A^m)} \{w \in \mathbb{R}^p, \langle w, v \rangle \le \mathcal{S}_A(v)\}$ were $\mathcal{V}(A^m)$ consists of m unit vectors and $\mathcal{S}_{A^m}(v) := \sup_{w \in A} \langle w, v \rangle$ is the support function of the convex A^m. We then define $A^{m,\varepsilon} := \bigcap_v \{w \in \mathbb{R}^p, \langle w, v \rangle \le \mathcal{S}_A(v) + \varepsilon\}$.) Then **Proposition 3.1** provides that for all $a, d > 0$

$$\rho_n(\mathcal{A}^{\mathrm{si}}(a,d)) \le K_6 \, \frac{B_n^{1/3} \, \log^{7/6}(p^d n^{d+1})}{b^{1/2} \, n^{1/6}} \quad \text{or} \quad \rho_n(\mathcal{A}^{\mathrm{si}}(a,d)) \le K_7 \, \frac{B_n^{2/3} \, \log(p^d n^{d+1})}{b^{1/2} \, n^{1/6}}. \tag{3.46}$$

whether condition (E.1) or (E.2) is satisfied.

We finally turn to the result we are interested in, that is **Proposition 3.2**. Let $\mathcal{A}^{\mathrm{sp}}$, be the class of s-sparsely convex sets in $\mathbb{R}^p$. Recall that $A = \bigcap_{q=1}^{Q} A_q$, where only s coordinate, say $J(A_q)$, are needed to define the set A_q. In step 1 we get the same bounds (3.46) but by replacing B_n by $s^3 B_n$. We can show that $d \le s(4s+1) \lesssim 4s^2$. In the first part of the second step of the proof we found that $\|\Sigma_s\|_{\mathcal{S}}^{1/2} \le s^{1/4}/b^{-1/2}$, where $\Sigma_s = \mathrm{Var}(\mathbf{X}_{t, s \in J(A_q)})$. In this same step, we found that the bound coming from the Berry-Essen theorem is $s^{5/2} B_n / n^{1/2}$. To conclude we just have to add them up. The second part of step 2 provides a similar bound as in step 1 and an additional use of Nazarov's inequality yields $d \log^{1/2}(pn) b^{-1/2} n^{-1}$, but this one is negligible. To conclude we get that

$$\rho_n(\mathcal{A}^{\mathrm{sp}}(s)) \le \; K_8 \frac{s^3 B_n^{1/3} \, \log^{7/6}(pn)}{b^{1/2} \, n^{1/6}} \quad \text{or} \quad \rho_n(\mathcal{A}^{\mathrm{sp}}(s)) \le \; K_9 \, \frac{s^3 B_n^{2/3} \, \log(pn)}{b^{1/2} \, n^{1/6}}, \tag{3.47}$$

whether condition (E.1) or (E.2) is satisfied.

3.4.4 Gumbel domain of attraction

In this section, we show that (3.28) is still true when $p = p_n \to \infty$. From the proof of Theorem 8 it is immediate that $\lambda_1^{-1}(M_n^p - b_n^p) \xrightarrow{d} \mathcal{G}$ provided that (3.28) tends to zero. In the following lemma we provide a simple sufficient condition for this.

Lemma 13. *Suppose that* (3.7) *holds. Then* (3.28) *tends to zero for any sequence* $p_n = O(n^{\gamma_0})$*, where* γ_0 *satisfies* (3.9)*.*

Proof of Lemma 13. Let us denote

$$a_{k,n} := \alpha_{k,p} \left(\frac{e^{-x}}{\alpha_{1,p}}\right)^{\lambda_1/\lambda_k} \left(\frac{2}{n}\right)^{\lambda_1/\lambda_k - 1}.$$

Let $p_n = O(n^{\gamma_0})$, for some $0 < \gamma_0 < 1$, we want to show that $\sum_{k=2}^{p_n} a_{k,n} \to 0$. We first remark that

$$\frac{1}{\alpha_{1,p}} = \prod_{j=2}^{p}(1 - \lambda_j/\lambda_1) \le (1 - \lambda_p/\lambda_1)^{p-1} \le 1. \tag{3.48}$$

From assumption (3.7), we know that the sequence $k\lambda_k$ is eventually decreasing. We can thus, choose an k_0 large enough such that λ_k

$$\frac{j\lambda_j}{k\lambda_k} \ge 1, \quad \text{for all} \quad 1 \le j \le k-1 \quad \text{and} \quad k \ge k_0.$$

Let then $k_1 \ge k_0$ be such that $\lambda_1/\lambda_k - 1 \ge k$, for all $k \ge k_1$. Then we denote

$$\underline{A}_n = \sum_{k=2}^{k_1-1} a_{k,n} \quad \text{and} \quad \overline{A}_n = \sum_{k=k_1}^{p_n} a_{k,n}. \tag{3.49}$$

For $k \ge k_1$ we have

$$\begin{aligned} |\alpha_{k,p}| &= \frac{1}{\prod_{j=1}^{k-1}(\lambda_j/\lambda_k - 1)\prod_{j=k+1}^{p}(1 - \lambda_j/\lambda_k)} \\ &\le \frac{1}{\prod_{j=1}^{k-1}\frac{k-j}{j}\prod_{j=k+1}^{p}\frac{j-k}{j}} = \binom{p}{k} \le \left(\frac{ep}{k}\right)^k. \end{aligned}$$

Using this, (3.48) and the fact that $\lambda_1/\lambda_k - 1 \ge k$, for $k \ge k_1$, we get that

$$\begin{aligned} |\overline{A}_n| &\le e^{-x}\sum_{k=k_1}^{p_n}\left(\frac{ep_n}{k}\right)^k\left(\frac{2e^{-x}}{n}\right)^{\lambda_1/\lambda_k - 1} \\ &\le e^{-x}\sum_{k=k_1}^{p_n}\left(\frac{2e^{1-x}}{k}\right)^k\left(\frac{p_n}{n}\right)^k \end{aligned}$$

for n large enough, i.e. such that $2e^{-x}/n < 1$. For $\gamma_0 < 1$, the general term of the last series converges to zero when $n \to \infty$, for all $k \ge k_1$, and is dominated by the summable term $\left(2e^{1-x}/k\right)^k$ we can thus conclude that $\overline{A}_n \to 0$.

Next, for $k < k_1$, set $\nu_k = \frac{1}{\prod_{j=1, j\neq k}^{k_1-1} |\lambda_j/\lambda_k - 1|}$. We have that

$$\begin{aligned}|\alpha_{k,p}| &\leq \frac{\nu_k}{\prod_{j=k_1}^{p}(1-\lambda_j/\lambda_k)} \\ &\leq \frac{\nu_k}{\prod_{j=k_1}^{p}\frac{j-k}{j}} \\ &= \nu_k \cdot \frac{p!(k_0-k-1)!}{(k_1-1)!(p-k)!} \\ &\leq \nu_k \cdot p^k \cdot \frac{(k_1-k-1)!}{(k_1-1)!}.\end{aligned}$$

Thus, we have

$$|\underline{A}_n| \leq \sum_{k=2}^{k_1} \nu_k \cdot p_n^k \, e^{-x\lambda_1/\lambda_k} \cdot \frac{(k_1-k-1)!}{(k_1-1)!} \left(\frac{2}{n}\right)^{\lambda_1/\lambda_k - 1} = O\left(\max_{2\leq k\leq k_1}\left\{\frac{p_n^k}{n^{\lambda_1/\lambda_k-1}}\right\}\right).$$

If we choose $\gamma_0 < \min_{2\leq k\leq k_1} \frac{1}{k}\left(\frac{\lambda_1}{\lambda_k} - 1\right)$ this term tends to zero as $n \to \infty$. □

3.4.5 Truncation

In this section, we show that when p is fixed, we do not need finite fourth order moments and can rather assume that $E\|X_1\|^s < \infty$, for some $s > 2$. In particular, we recover the scalar result of Davis and Mikosch (1999). To this end, we consider the truncated array

$$\widetilde{X}_t = X_t \mathbb{1}_{\|X_t\|\leq n^{1/s}} - E[X_t \mathbb{1}_{\|X_t\|\leq n^{1/s}}]. \tag{3.50}$$

Thus, for all $n \geq 1$, $\widetilde{X}_t$ is bounded. The next Lemma ensures that we can replace $M_n^{X,p}$ by $M_n^{\widetilde{X},p}$ in the proof of Theorem 8.

Lemma 14. *Suppose $E\|X_1\|^s < \infty$, for $s > 2$, then we have that*

$$M_n^{X,p} - M_n^{\widetilde{X},p} \underset{n\to\infty}{\longrightarrow} 0 \; a.s.$$

Proof of Lemma 14. The proof of Lemma 3.3 in Davis and Mikosch (1999) can be easily adapted to our setting. □

We can define similarly $\widetilde{Y}_t$ and $M_n^{\widetilde{Y},p}$, where as usual, $Y_1, \ldots, Y_n$ are i.i.d. Gaussian elements with same variance as X_t. It is obvious that Lemma 14 applies as well to $\widetilde{Y}_t$. We conclude this section with the analogue of Proposition 2 for

$$\widetilde{\rho}_{n,p} = \sup_{x\in\mathbb{R}} |P(M_n^{\widetilde{X},p} \leq x) - P(M_n^{\widetilde{Y},p} \leq x)|. \tag{3.51}$$

Proposition 3. *Let $\widetilde{\rho}_{n,p}$ be given as in* (3.51)*. There exists a universal constant K' such that*

$$\widetilde{\rho}_{n,p} \leq K' \cdot B_n^{1/3} \, \frac{p^3 \, \log^{7/6}(pn^2)}{\widetilde{\lambda}_p^{1/2} \, n^{1/6}}, \tag{3.52}$$

where $B_n = n^{1/s}$.

Corollary 2. *Let $p \geq 1$ be a fixed integer. If $E\|X_1\|^s < \infty$, for some $s > 2$ and $\lambda_p > 0$, then we have that $\widetilde{\rho}_{n,p} \to 0$, when $n \to +\infty$.*

Proof of Proposition 3. We proceed as in Proposition 2 except that we consider the $2pq$-variate element $\widetilde{\mathbf{X}}_t = \widetilde{\xi}_t^{(p)} \otimes W_t^{(2q)}$ where $\widetilde{\xi}_t^{(p)} = (\langle \widetilde{X}_t, \tilde{v}_1 \rangle, \ldots, \langle \widetilde{X}_t, \tilde{v}_p \rangle)'$, and $\tilde{v}_j$ denotes the eigenfunctions associated to the ordered eigenvalues $\widetilde{\lambda}_j$ of $\mathrm{Var}(\widetilde{X}_1)$. We also rather use the first bound in (3.47). We thus need to verify conditions (M.1) (M.2) and (E.1). By applying Lemma 11 to $\widetilde{\mathbf{X}}_t$ we deduce that Condition (M.1") holds with $b = \widetilde{\lambda}_p/2$. Now suppose that $k = 1, 2$ and $2 \leq s < 3$. Since the series

$$\sum_{i=1}^{\infty} t^{-(2+k)/s} E|\mathbf{X}_{tj}|^{2+k} \mathbb{1}_{\|\mathbf{X}_t\| \leq t^{1/s}}$$

converges provided that $E|X_{ij}|^s < \infty$ for each $t \geq 1$ and $j = 1, \ldots, d$, we have that

$$n^{-(2+k)/s} \sum_{t=1}^{n} E|\mathbf{X}_{tj}|^{2+k} \mathbb{1}_{\|\mathbf{X}_t\| \leq t^{1/s}} \to 0$$

as $n \to \infty$ using Kronecker's lemma. Hence,

$$n^{-1} \sum_{t=1}^{n} E|\mathbf{X}_{tj}|^{2+k} \mathbb{1}_{\|\mathbf{X}_{tj}\| \leq t^{1/s}} = o(n^{(2+k)/s-1})$$

as $n \to \infty$. Let us observe that $(2+k)/s - 1 < k/2$ if $s > 2$, and thus condition (M.2) holds for $B_n = n^{1/s}$. For condition (E.1), note that $|\widetilde{\mathbf{X}}_{t,j}|/B_n \leq t^{1/s}/n^{1/s} \leq 1$. To conclude it suffices to apply the first bound in (3.47) to this setting. $\square$

Proof of Corollary 2. From the definition (3.50) we have that

$$\mathrm{Var}(\widetilde{X}_t) = EX_t \otimes X_t \mathbb{1}_{\|X_t\| \leq n^{1/s}} \to EX_t \otimes X_t = \mathrm{Var}(X_t) \quad (n \to \infty).$$

Therefore, $\widetilde{\lambda}_p \to \lambda_p > 0$ and it suffices to apply Proposition 3 with $s > 2$. $\square$

3.4.6 Finite dimensional setting

Proof of Theorem 11. Let p be a fixed integer and define the rescaled innovations $\widetilde{\mathbf{Z}}_t = \Sigma^{-1/2}\mathbf{Z}_t$. It holds that

$$\left| P\big(M_n^{\widetilde{\mathbf{Z}}} - c_n^p \leq x\big) - e^{-e^{-x}} \right| \leq \rho_{n,p} + \left| P\big(M_n^{\mathbf{Y}} - c_n^p \leq x\big) - e^{-e^{-x}} \right|, \tag{3.53}$$

where $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ are i.i.d. $\mathcal{N}_p(0, I_p)$. Furthermore, we have that

$$\left\{ \left\| \mathcal{X}_n^{\mathbf{Y}}(\omega_j) \right\|^2,\ j = 1, \ldots, q \right\} \stackrel{d}{=} \left\{ \zeta_j^{(p)},\ j = 1, \ldots, q \right\},$$

where the $\zeta_j^{(p)}$ are i.i.d. Erlang$(p, 1)$. Their cumulative distribution function is

$$F^{(p)}(x) = 1 - \sum_{k=0}^{p-1} \frac{x^k}{k!} e^{-x}.$$

Then, with the definition of c_n^p we get

$$P\left(M_n^{\mathbf{Y}} - c_n^p \leq x\right) = \left(1 - e^{-x}\frac{2}{n}\frac{(p-1)!}{\log^{p-1}(n/2)}\sum_{k=0}^{p-1}\frac{(x+c_n^p)^k}{k!}\right)^q.$$

By a routine argument one can show that this converges to the Gumbel distribution function for every given $x \in \mathbb{R}$ and $p \geq 1$, see e.g. in Kang and Serfolo (1999). Now the result follows from Lemma 15 below. □

The next lemma is a generalization of Theorem 3 in Walker (1965) for multivariate linear processes, $\boldsymbol{X}_t = \sum_{k=0}^{\infty}\Psi_k \mathbf{Z}_{t-k}$, with i.i.d. innovations $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$ in $\mathbb{R}^p$.

Lemma 15. *Under Assumption 6, we have that*

$$\left|L_n^{\mathbf{X}} - M_n^{\Sigma^{-1/2}\mathbf{Z}}\right| \xrightarrow{P} 0. \tag{3.54}$$

Proof of Lemma 15. First remark that $\mathbf{F}_\omega = \Psi_\omega \Sigma \Psi_\omega^*$, where $\Psi_\omega = \sum_{k\geq 0}\Psi_k e^{\mathrm{i}k\omega}$ and where A^* is the conjugate transpose of a matrix A, and thus, we also have that $A \otimes B = AB^*$. Also recall that $\widetilde{\mathbf{Z}}_t = \Sigma^{-1/2}\mathbf{Z}_t$.

Next let us first notice that there exists a constant $\kappa < \infty$, such that

$$E \max_{1\leq j\leq q}\left\|\sum_{t=1}^{r}\mathbf{Z}_t e^{-\mathrm{i}t\omega_j}\right\|^2 \leq \kappa r^{3/2}. \tag{3.55}$$

This can be shown by first noting that

$$\left\|\sum_{t=1}^{r}\mathbf{Z}_t e^{-\mathrm{i}t\omega_j}\right\|^2 = \sum_{|h|<r}\sum_{t=1}^{r-|h|}\langle \mathbf{Z}_{t+|h|}, \mathbf{Z}_t\rangle e^{-ih\omega_j} \leq \sum_{|h|<r}\left|\sum_{t=1}^{r-|h|}\langle \mathbf{Z}_{t+|h|}, \mathbf{Z}_t\rangle\right|.$$

Then, applying the Lyapunov's inequality we get

$$E\left|\sum_{t=1}^{r-|h|}\langle \mathbf{Z}_{t+|h|}, \mathbf{Z}_t\rangle\right| \leq \left(\sum_{t=1}^{r-|h|}\sum_{s=1}^{r-|h|}E\langle \mathbf{Z}_{t+|h|}, \mathbf{Z}_t\rangle\langle \mathbf{Z}_{s+|h|}, \mathbf{Z}_s\rangle\right)^{1/2} \leq \left(2(r-|h|)E\|\mathbf{Z}_1\|^4\right)^{1/2}.$$

Here we used the independence of the $\mathbf{Z}_t$'s. Finally, it holds that $\sum_{|h|<r}(r-|h|)^{1/2} \leq 2r^{3/2}$.

Of course, (3.55) holds analogously with $\mathbf{Z}_t$ replaced by $\widetilde{\mathbf{Z}}_t$. Moreover, we can even extend this bound for the Fourier transform of $\mathbf{X}_t$:

$$
\begin{aligned}
E\max_{1\le j\le q}\left\|\sum_{t=1}^{n}\mathbf{X}_t e^{-\mathrm{i}t\omega_j}\right\|^2 &= E\max_{1\le j\le q}\left\|\sum_{k=0}^{\infty}\Psi_k\left(\sum_{s=1-k}^{n-k}\mathbf{Z}_s e^{-\mathrm{i}s\omega_j}\right)e^{-\mathrm{i}k\omega_j}\right\|^2\\
&\le E\max_{1\le j\le q}\left(\sum_{k=0}^{\infty}\|\Psi_k\|\cdot\left\|\sum_{t=1-k}^{n-k}\mathbf{Z}_t e^{-\mathrm{i}t\omega_j}\right\|\right)^2\\
&\le \sum_{k,\ell=0}^{\infty}\|\Psi_k\|\,\|\Psi_\ell\|\cdot E\Big[\max_{1\le j\le q}\left\|\sum_{t=1-k}^{n-k}\mathbf{Z}_t e^{-\mathrm{i}t\omega_j}\right\|\,\left\|\sum_{t=1-\ell}^{n-\ell}\mathbf{Z}_t e^{-\mathrm{i}t\omega_j}\right\|\Big]\\
&\le \left(\sum_{k=0}^{\infty}\|\Psi_k\|\right)^2\cdot E\max_{1\le j\le q}\left\|\sum_{t=1-k}^{n-k}\mathbf{Z}_t e^{-\mathrm{i}t\omega_j}\right\|^2\\
&\le \left(\sum_{k=0}^{\infty}\|\Psi_k\|\right)^2\cdot E\max_{1\le j\le q}\left\|\sum_{t=1}^{n}\mathbf{Z}_t e^{-\mathrm{i}t\omega_j}\right\|^2\le \kappa n^{3/2}.
\end{aligned}
$$

Next set $\widetilde{L}_n^{\mathbf{X}} := \max_{1\le j\le q}\operatorname{tr}\Big(\mathbf{F}_{\omega_j}^{-1}I_n^{\mathbf{X}}(\omega_j)\Big)$ and $A_\omega=\Psi_\omega\Sigma^{1/2}$. Using $\operatorname{tr}(AB)\le \|A\|_{\mathcal{S}}\|B\|_{\mathcal{S}}$ we infer

$$
|\operatorname{tr}(\widetilde{L}_n^{\mathbf{X}}-L_n^{\mathbf{X}})|\le \max_{1\le j\le q}\|\mathbf{F}_{\omega_j}^{-1}-\widehat{\mathbf{F}}_{\omega_j}^{-1}\|_{\mathcal{S}}\times\max_{1\le j\le q}\|I_n^{X}(\omega_j)\|_{\mathcal{S}}.
$$

With elementary matrix algebra we deduce that

$$
\mathbf{F}_\omega^{-1}-\widehat{\mathbf{F}}_\omega^{-1}=\mathbf{F}_\omega^{-1}(\widehat{\mathbf{F}}_\omega-\mathbf{F}_\omega)\Big(I+\mathbf{F}_\omega^{-1}(\widehat{\mathbf{F}}_\omega-\mathbf{F}_\omega)\Big)^{-1}\mathbf{F}_\omega^{-1},
$$

and thus

$$
\|\mathbf{F}_\omega^{-1}-\widehat{\mathbf{F}}_\omega^{-1}\|_{\mathcal{S}}\le\|\mathbf{F}_\omega^{-1}\|^2\|\mathbf{F}_\omega-\widehat{\mathbf{F}}_\omega\|\Big\|\Big(I+\mathbf{F}_\omega^{-1}(\widehat{\mathbf{F}}_\omega-\mathbf{F}_\omega)\Big)^{-1}\Big\|_{\mathcal{S}}.
$$

From Assumptions (3.21) and (3.20) we infer that

$$
\|\mathbf{F}_{\omega_j}^{-1}-\widehat{\mathbf{F}}_{\omega_j}^{-1}\|_{\mathcal{S}}=o_P(n^{-1/4})\times\Big\|\Big(I+\mathbf{F}_{\omega_j}^{-1}(\widehat{\mathbf{F}}_{\omega_j}-\mathbf{F}_{\omega_j})\Big)^{-1}\Big\|_{\mathcal{S}}.
$$

It is elementary that

$$
\Big\|\Big(I+\mathbf{F}_\omega^{-1}(\widehat{\mathbf{F}}_\omega^{-1}-\mathbf{F}_\omega^{-1})\Big)^{-1}\Big\|_{\mathcal{S}}\le\sum_{k\ge0}\Big(\|\mathbf{F}_\omega^{-1}\|\|\widehat{\mathbf{F}}_\omega-\mathbf{F}_\omega\|_{\mathcal{S}}\Big)^k.
$$

By our assumptions $\max_{1\le j\le q}\|\widehat{\mathbf{F}}_{\omega_j}-\mathbf{F}_{\omega_j}\|_{\mathcal{S}}\xrightarrow{P}0$, thus we conclude that $\max_{1\le j\le q}\|\mathbf{F}_{\omega_j}^{-1}-\widehat{\mathbf{F}}_{\omega_j}^{-1}\|_{\mathcal{S}}=o_P(n^{-1/4})$.

It remains then to show that $|\widetilde{L}_n^{\mathbf{X}}-M_n^{\widetilde{\mathbf{Z}}}|\xrightarrow{P}0$. To this end we notice that

$$
\begin{aligned}
|\widetilde{L}_n^{\mathbf{X}}-M_n^{\widetilde{\mathbf{Z}}}| &\le \max_{1\le j\le q}\left|\operatorname{tr}\Big(\mathbf{F}_{\omega_j}^{-1}I_n^{\mathbf{X}}(\omega_j)-I_n^{\widetilde{\mathbf{Z}}}(\omega_j)\Big)\right|\\
&=\max_{1\le j\le q}\left|\operatorname{tr}\Big((A_{\omega_j}A_{\omega_j}^*)^{-1}I_n^{\mathbf{X}}(\omega_j)-I_n^{\widetilde{\mathbf{Z}}}(\omega_j)\Big)\right|\\
&=\max_{1\le j\le q}\left|\operatorname{tr}\Big(A_{\omega_j}^{-1}\Big(I_n^{\mathbf{X}}(\omega_j)-A_{\omega_j}I_n^{\widetilde{\mathbf{Z}}}(\omega_j)A_{\omega_j}^*\Big)(A_{\omega_j}^*)^{-1}\Big)\right|\\
&\le\max_{1\le j\le q}\|\mathbf{F}_{\omega_j}^{-1}\|_{\mathcal{S}}\times\max_{1\le j\le q}\Big\|I_n^{\mathbf{X}}(\omega_j)-A_{\omega_j}I_n^{\widetilde{\mathbf{Z}}}(\omega_j)A_{\omega_j}^*\Big\|_{\mathcal{S}}.
\end{aligned}
$$

In consequence of (3.20) it remains to show that

$$\max_{1\le j\le q}\left\|I_n^{\mathbf{X}}(\omega_j)-A_{\omega_j}I_n^{\widetilde{\mathbf{Z}}}(\omega_j)A_{\omega_j}^*\right\|_{\mathcal{S}}\xrightarrow{P}0. \tag{3.56}$$

We have

$$\begin{aligned}
I_n^{\mathbf{X}}(\omega_j)-A_{\omega_j}I_n^{\widetilde{\mathbf{Z}}}(\omega_j)A_{\omega_j}^* &= \left(\mathcal{X}_n^{\mathbf{X}}(\omega_j)-A_{\omega_j}\mathcal{X}_n^{\widetilde{\mathbf{Z}}}(\omega_j)\right)\left(\mathcal{X}_n^{\mathbf{X}}(\omega_j)-A_{\omega_j}\mathcal{X}_n^{\widetilde{\mathbf{Z}}}(\omega_j)\right)^*\\
&\quad+\left(\mathcal{X}_n^{\mathbf{X}}(\omega_j)-A_{\omega_j}\mathcal{X}^{\widetilde{\mathbf{Z}}}(\omega_j)\right)\left(A_{\omega_j}\mathcal{X}_n^{\widetilde{\mathbf{Z}}}(\omega_j)\right)^*\\
&\quad+A_{\omega_j}\mathcal{X}_n^{\widetilde{\mathbf{Z}}}(\omega_j)\left(\mathcal{X}_n^{\mathbf{X}}(\omega_j)-A_{\omega_j}\mathcal{X}_n^{\widetilde{\mathbf{Z}}}(\omega_j)\right)^*.
\end{aligned}$$

Hence we deduce that

$$\begin{aligned}
\max_{1\le j\le q}\left\|I_n^{\mathbf{X}}(\omega_j)-A_{\omega_j}I_n^{\widetilde{\mathbf{Z}}}(\omega_j)A_{\omega_j}^*\right\|_{\mathcal{S}} &\le \max_{1\le j\le q}\left\|\mathcal{X}_n^{\mathbf{X}}(\omega_j)-A_{\omega_j}\mathcal{X}_n^{\widetilde{\mathbf{Z}}}(\omega_j)\right\|_{\mathcal{S}}^2\\
&+2\max_{1\le j\le q}\left\|A_{\omega_j}\mathcal{X}_n^{\widetilde{\mathbf{Z}}}(\omega_j)\right\|_{\mathcal{S}}\cdot\left\|\mathcal{X}_n^{\mathbf{X}}(\omega_j)-A_{\omega_j}\mathcal{X}_n^{\widetilde{\mathbf{Z}}}(\omega_j)\right\|_{\mathcal{S}}.
\end{aligned}$$

Note that from $A_\omega\mathcal{X}_n^{\widetilde{\mathbf{Z}}}(\omega)=\Psi_\omega\mathcal{X}_n^{\mathbf{Z}}(\omega)$ it follows that

$$\mathcal{X}_n^{\mathbf{X}}(\omega)-A_\omega\mathcal{X}_n^{\widetilde{\mathbf{Z}}}(\omega)=n^{-1/2}\sum_{k=0}^{\infty}\Psi_k\Delta_{n,k}(\omega)e^{-\mathrm{i}k\omega}, \tag{3.57}$$

where $\Delta_{n,k}(\omega)=\sum_{s=1-k}^{n-k}\mathbf{Z}_se^{-\mathrm{i}s\omega}-\sum_{t=1}^{n}\mathbf{Z}_te^{-\mathrm{i}t\omega}$. Then, using (3.57) and (3.55) for $r=\min\{k,n\}$ we deduce that $E\max_{1\le j\le q}\|\Delta_{n,k}(\omega_j)\|\le 2\kappa^{1/2}\,(k\wedge n)^{3/4}$. We then deduce that

$$\begin{aligned}
E\max_{1\le j\le q}\left\|\mathcal{X}_n^{\mathbf{X}}(\omega_j)-A_{\omega_j}\mathcal{X}_n^{\widetilde{\mathbf{Z}}}(\omega_j)\right\| &\le n^{-1/2}\sum_{k=0}^{\infty}\|\Psi_k\|\,E\max_{1\le j\le q}\|\Delta_{n,k}(\omega)\|\\
&\le 2\kappa^{1/2}n^{-1/2}\left(\sum_{k=0}^{n-1}k^{3/4}\|\Psi_k\|+n^{3/4}\sum_{k\ge n}\|\Psi_k\|\right)\\
&\le 2\kappa^{1/2}n^{-1/4}\left(\sum_{k=0}^{n-1}k^{1/2}(k/n)^{1/4}\|\Psi_k\|+\sum_{k\ge n}k^{1/2}\|\Psi_k\|\right)\\
&= o(n^{-1/4}).
\end{aligned}$$

Finally, from $\|\Psi_\omega\mathcal{X}_n^{\mathbf{Z}}(\omega)\|\le\|\Psi_\omega\|\cdot\|\mathcal{X}_n^{\mathbf{Z}}(\omega)\|$ and (3.55), we get that

$$\begin{aligned}
E\max_{1\le j\le q}\left\|A_{\omega_j}\mathcal{X}_n^{\widetilde{\mathbf{Z}}}(\omega_j)\right\| &\le \max_{1\le j\le q}\left\|\Psi_{\omega_j}\right\|n^{-1/2}\left(E\max_{1\le j\le q}\|\sum_{t=1}^{n}\mathbf{Z}_te^{-\mathrm{i}t\omega_j}\|^2\right)^{1/2}\\
&\le \max_{1\le j\le q}\left\|\Psi_{\omega_j}\right\|2\kappa^{1/2}\;n^{1/4}.
\end{aligned}$$

To conclude, recall that all norms are equivalent in finite dimension. □

4 Functional GARCH models: the quasi-likelihood approach and its applications

Clément Cerovecki*, Christian Francq†,
Siegfried Hörmann‡ and Jean-Michel Zakoïan§

Abstract

The increasing availability of high frequency data has initiated many new research areas in statistics. Functional data analysis (FDA) is one of these disciplines. In FDA, densely observed data are transformed into curves and then each (random) curve is considered as one data object. A natural, but still relatively unexplored, context for FDA methods is related to financial data, where high-frequency trading currently takes a significant proportion of trading volumes. Recently, articles on functional versions of the famous ARCH and GARCH models have appeared. Due to their technical complexity, existing estimators of the underlying functional parameters are moment based—an approach which is known to be relatively inefficient in this context. In this paper, we promote an alternative quasi-likelihood approach, for which we derive consistency and asymptotic normality results. We support the relevance of our approach by simulations and illustrate its use by forecasting realised volatility of the S&P100 Index.

JEL Classification: C13, C32 and C58.

Keywords: Functional time series, High-frequency volatility models, Intraday returns, Functional QMLE, Stationarity of functional GARCH.

4.1 Introduction

Financial time series modelling is of great importance to monitor the evolution of prices, stock indexes or exchange rates and to predict future developments, such as the risk associated to certain asset allocations. Risk is very much related to the volatility of the financial process and, hence, models for volatility are of special importance. A milestone in volatility modelling has been set by Engle (1982), with the introduction of the now-famous and widely-used ARCH model. Many extensions followed this groundbreaking work, most notably the GARCH model by Bollerslev (1986) which allows a more parsimonious fit in comparison to ARCH processes. The success of these

*Université libre de Bruxelles, Boulevard du Triomphe, CP210 B-1050 Brussels - Belgium Office: NO O.9.211. *E-Mail address*: clement.cerovecki@ulb.ac.be

†CREST and University of Lille, BP 60149, 59653 Villeneuve d'Ascq cedex, France. *E-Mail address*: christian.francq@univ-lille3.fr

‡Université libre de Bruxelles, Boulevard du Triomphe, CP210 B-1050 Brussels - Belgium Office: NO O.9.115

§Corresponding author: Jean-Michel Zakoïan, University of Lille and CREST, 5 Avenue Henry Le Chatelier, 91120 Palaiseau, France. *E-Mail address*: zakoian@ensae.fr, Phone number: 33.1.41.17.77.25.

models is founded on their mathematical feasibility and on their ability to feature many of the *stylized facts* that researchers have been observing in empirical investigations of financial data. In particular, the models are able to capture a non-constant conditional variance of time series. For details on GARCH models, see e.g. Francq and Zakoian (2011) and the references therein.

In practical applications, GARCH models and their variations are adequate for daily or weekly return data. But, due to the availability of high-frequency financial time series and their importance for the financial industry, it is desirable to provide corresponding models and adequate statistical methodology for data that are given at a higher resolution. In this paper, we adopt the theory of *functional time series* to approach this challenge. A functional time series is a sequence of observations $(X_t\colon 1 \leq t \leq n)$, where each random object X_t is a curve $(X_t(u)\colon u \in [0,1])$. The interval $[0,1]$ is chosen for convenience and does not impose any restriction on generality. In our context it represents intraday time. For example, $X_t(u)$ might denote the price of an asset on day t at intraday time u. If we then consider the log-returns $y_t(u) = \log X_t(u) - \log X_t(u-\tau)$ on some τ-interval or the intraday log-increments $\tilde{y}_t(u) = \log X_t(u) - \log X_t(0)$, then it seems plausible that such common transformations yield stationary functional processes (as processes in the discrete time t), in which case a variety of tools can be employed for inference on the intraday pattern.

Functional time series methods have received increasing attention during the past few years. To give a small sample of some very recent articles with many further references we refer to the following papers: Hörmann and Kokoszka (2010) and Eichler and van Delft (2017) for structural results, Horváth et al. (2014) and Aue and Klepsch (2017a) for inferential procedures, Paparoditis (2017) and Zhu and Politis (2017) for functional time series bootstrapping methods and Aue et al. (2015) and Klepsch and Klüppelberg (2017)) for forecasting algorithms.

In this paper, we consider some adequate functional models to describe, for instance, the functional time series (y_t) or $(\tilde{y}_t)$ as defined above. The first attempt to generalise GARCH models to *functional time series* was made in Hörmann et al. (2013), where a functional version of the ARCH(1) was proposed. Later, this model was extended in Aue et al. (2016) to a functional GARCH(1,1). Both models rely on recurrence equations with unknown operators and curves. As for the estimation of these quantities, Hörmann et al. (2013) proposed a moment-based estimator and showed its consistency. Their approach allows to deal with a fully functional (and potentially infinite-dimensional) parameter space. The situation is more complicated in the GARCH context. Aue et al. (2016) proposed a least squares estimator based on the recursive empirical volatility. This approach comes at a price: the authors have to reduce the functional model to a multivariate model via some dimension reduction to a fixed finite dimension. Moreover, it is know from scalar GARCH theory that the least-squares estimators lack efficiency.

In this paper, *we propose an estimator inspired by the classical GARCH QML (Quasi-Maximum Likelihood) method* (Section 4.3). The definition of a QMLE is far from straightforward in that context, because a likelihood cannot be written for curves. Our estimator is based on the projection of the squared process onto a set of non-negative valued instrumental functions. We give regularity conditions for *consistency and asymptotic normality.* As a side result, we obtain the consistency and asymptotic normality of a semi-strong (i.e. with non-iid innovations) multivariate CCC-GARCH. We also obtain a sufficient condition for existence of stationary functional GARCH processes (Section 4.2.2) which generalises Aue et al. (2016). We use the top Lyapunov exponent formulation, and our condition is very similar to the sufficient and necessary condition that can be obtained in the finite dimensional case. Our results also extend to higher order models, i.e. functional GARCH(p, q). In terms of application, we use our model to *predict realised volatility* which is an important risk measure (Section 4.5.2).

The rest of the paper is organised as follows: in Section 4.2, we introduce the model equations, some notations and discuss the stationarity. In Section 4.3, we introduce our estimation procedure and detail its asymptotic properties. In Section 4.4, we extend our consistency results to infinite-dimensional models. The subsequent sections deal with practical aspects of the implementation and some empirical illustrations which demonstrate the superiority of the QMLE compared to existing methods. Technical results and most proofs are given in the Appendix.

4.2 Functional GARCH(p,q) model

4.2.1 Preliminaries

For convenience we first review notation. We denote by H the Hilbert space of square integrable functions with domain $[0, 1]$. It will serve as the basic space on which the functional observation, that is considered in this paper, is defined. The Hilbert space is equipped with inner product $\langle \cdot, \cdot \rangle$ and the resulting norm $\| \cdot \|$. If x and y are both functions of H (respectively, vectors of $\mathbb{R}^d$), then we denote by xy their point-wise (resp. component-wise) product. We denote by $\mathcal{L}(H)$ the space of bounded linear operators on H and use bold notation for its elements. Hence, for $\boldsymbol{\alpha} \in \mathcal{L}(H)$, $x \in H$ and $u \in [0, 1]$ we have that $\boldsymbol{\alpha}(x)$ is the image in H of $\boldsymbol{\alpha}$ applied to x, whereas $x(u)$ is the real-valued image of the function x evaluated at u. Moreover, we use the standard convention for combining operators, i.e. that $\boldsymbol{\alpha\beta} := \boldsymbol{\alpha} \circ \boldsymbol{\beta}$ and $\boldsymbol{\alpha}^2 := \boldsymbol{\alpha} \circ \boldsymbol{\alpha}$ for $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathcal{L}(H)$. We recall that $\mathcal{L}(H)$, equipped with the usual operator norm $\|\boldsymbol{\alpha}\| := \sup_{\|x\| \leq 1} \|\boldsymbol{\alpha}(x)\|$, is a Banach space. This norm is sub-multiplicative, i.e. $\|\boldsymbol{\alpha\beta}\| \leq \|\boldsymbol{\alpha}\| \, \|\boldsymbol{\beta}\|$. In some places we also make use of the supremum norm: $\|x\|_\infty = \inf\{a > 0,\ |x(u)| < a, \text{ for } \lambda\text{–almost every } u \in [0, 1]\}$. For $x, y \in H$, we define the operator $x \otimes y := x\langle \cdot, y \rangle$.

We define the subspaces $H^+ = \{x \in H : x(u) \geq 0, \text{ for almost every } u \in [0, 1]\}$ and $H^+_* = \{x \in H : x(u) > 0, \text{ for almost every } u \in [0, 1]\}$. Let $\mathcal{K}(H)$ denote the space of kernel operators on H, i.e. if $\boldsymbol{\alpha} \in \mathcal{K}(H)$ then there exists a function $K_{\boldsymbol{\alpha}} : [0, 1] \times [0, 1] \to \mathbb{R}$ such that $\boldsymbol{\alpha}(x)(u) = \int K_{\boldsymbol{\alpha}}(u, v)x(v)dv$. For simplicity, we will often write $\int$ instead of $\int_0^1$. Let $\mathcal{L}^+(H)$ denote the space of operators which map H^+ onto H^+ and note that an operator $\boldsymbol{\alpha} \in \mathcal{K}^+(H) := \mathcal{L}^+(H) \cap \mathcal{K}(H)$, if and only if its kernel $K_{\boldsymbol{\alpha}}(\cdot, \cdot)$ is non-negative.

For any integer $k \geq 2$, the product space $H^k = H \times \cdots \times H$ naturally inherits the Hilbert space structure by defining its scalar product as $\langle x, y \rangle = \sum_{i=1}^k \langle x_i, y_i \rangle$, for $x, y \in H^k$. In this context, it will be useful to represent elements and operators as k-dimensional vectors with values in H and $k \times k$ matrices with values in $\mathcal{L}(H)$, respectively. For example, if $k = 2$, we consider the operator

$$\boldsymbol{\alpha} : x \longmapsto \begin{pmatrix} \boldsymbol{\alpha}_{11} & \boldsymbol{\alpha}_{12} \\ \boldsymbol{\alpha}_{21} & \boldsymbol{\alpha}_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} := \begin{pmatrix} \boldsymbol{\alpha}_{11}(x_1) + \boldsymbol{\alpha}_{12}(x_2) \\ \boldsymbol{\alpha}_{21}(x_1) + \boldsymbol{\alpha}_{22}(x_2) \end{pmatrix}$$

where $x = (x_1, x_2)^\top \in H^2$ and $\boldsymbol{\alpha}_{11}$, $\boldsymbol{\alpha}_{12}$, $\boldsymbol{\alpha}_{21}$ and $\boldsymbol{\alpha}_{22} \in \mathcal{L}(H)$.

We are now ready to introduce our general model.

Definition 1. *Let $(\eta_t)_{t\in\mathbb{Z}}$ be a sequence of i.i.d. random elements of H. A functional GARCH(p,q) process $(y_t)_{t\in\mathbb{Z}}$ is defined as a stationary solution of the equations*

$$y_t = \sigma_t \eta_t, \tag{4.1}$$

$$\sigma_t^2 = \delta + \sum_{i=1}^{q} \boldsymbol{\alpha}_i (y_{t-i}^2) + \sum_{j=1}^{p} \boldsymbol{\beta}_j (\sigma_{t-j}^2), \tag{4.2}$$

where $\delta \in H_*^+$ and $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_q, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p \in \mathcal{K}^+(H)$. Such a solution is called non-anticipative if $\sigma_t = \sigma(\eta_{t-1}, \eta_{t-2}, \dots)$ for some measurable function σ.

Under the assumption that $E(\eta_t(u)) = 0$ and $E(\eta_t^2(u)) = 1$, the variable $\sigma_t^2(u)$ can be interpreted as the volatility at day t and intraday time u, i.e. the variance of the return $y_t(u)$ conditional upon the sigma algebra $\mathcal{F}_t$ generated by $(\eta_s)_{s\le t}$. Note that this volatility may depend on all past returns, not only on those corresponding to intraday time u of the previous days. For instance, let $p = 0, q = 1$ (ARCH(1)), and suppose that $\boldsymbol{\alpha}$ is a kernel operator, with constant kernel $K_{\boldsymbol{\alpha}}(u, v) := a$, then the pattern of the intraday volatility $\sigma_t^2(u) = \delta(u) + a \int y_{t-1}^2(v)dv$ is essentially given by that of δ. Moreover, it depends on the previous day through the so-called 'integrated volatility' given by the integral. If now, the kernel has the form $K_{\boldsymbol{\alpha}}(u, v) = a\phi(u - v)$, where ϕ denotes a density function with mode at zero, we have $\sigma_t^2(u) = \delta(u) + a \int \phi(u - v) y_{t-1}^2(v)dv$ and the volatility of intraday-time u is mainly driven by the volatility of the previous day 'around' time u. It is, thus, clear that Model (4.1) allows for a great flexibility through the choice of the operators $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and the pattern of the intercept δ.

A key feature of the GARCH model is that it captures well the dynamics of volatility observed in financial data. In our *functional* setting, we propose the following interpretation of the volatility curves. For any fixed $u \in [0, 1]$, we have that

$$P\left(|y_t(u)| < c \mid \mathcal{F}_{t-1}\right) = P\left(|\eta_t(u)| < c/\sigma_t(u) \mid \mathcal{F}_{t-1}\right) = 1 - \alpha, \tag{4.3}$$

if we take $c = \sigma_t(u) \cdot Q_{1-\alpha/2}^{\eta(u)}$. Consider, for example, a process with Gaussian innovations (η_t) such that $\mathrm{Var}(\eta_t(u)) = 1$ for all $u \in [0, 1]$. We can then interpret the region $\{[-2\sigma_t(u), 2\sigma_t(u)] \colon u \in [0, 1]\}$ as the prediction interval of $y_t(u)$ at (approximate) level $\alpha = .05$. We show these curves and their estimation in Figure 9 (in a setting that will be described below) at two different scales: 7 and 100 days, respectively. The data generating process is described in Example 2. On the first figure we observe the sensibility to shocks of the volatility curves. On the second figure we can observe the persistence of the volatility curves on a larger scale.

One interest of the functional GARCH model is that it allow for prediction of the next day's volatility curve. At the end of day $t - 1$, the whole volatility curve of day t can be predicted. It is, thus, possible to predict the realised volatility $\sum_{j=1}^{\lfloor 1/\tau \rfloor} y_t^2(j\tau)$ for some given time unit $\tau \in (0, 1)$, or any other realised measure of volatility. This will be illustrated in Section 4.5.2.

4.2.2 Existence of stationary solutions

In light of Definition 1, an evident question concerns the existence of a strictly stationary and non-anticipative solution to the functional GARCH equations. To respond to this

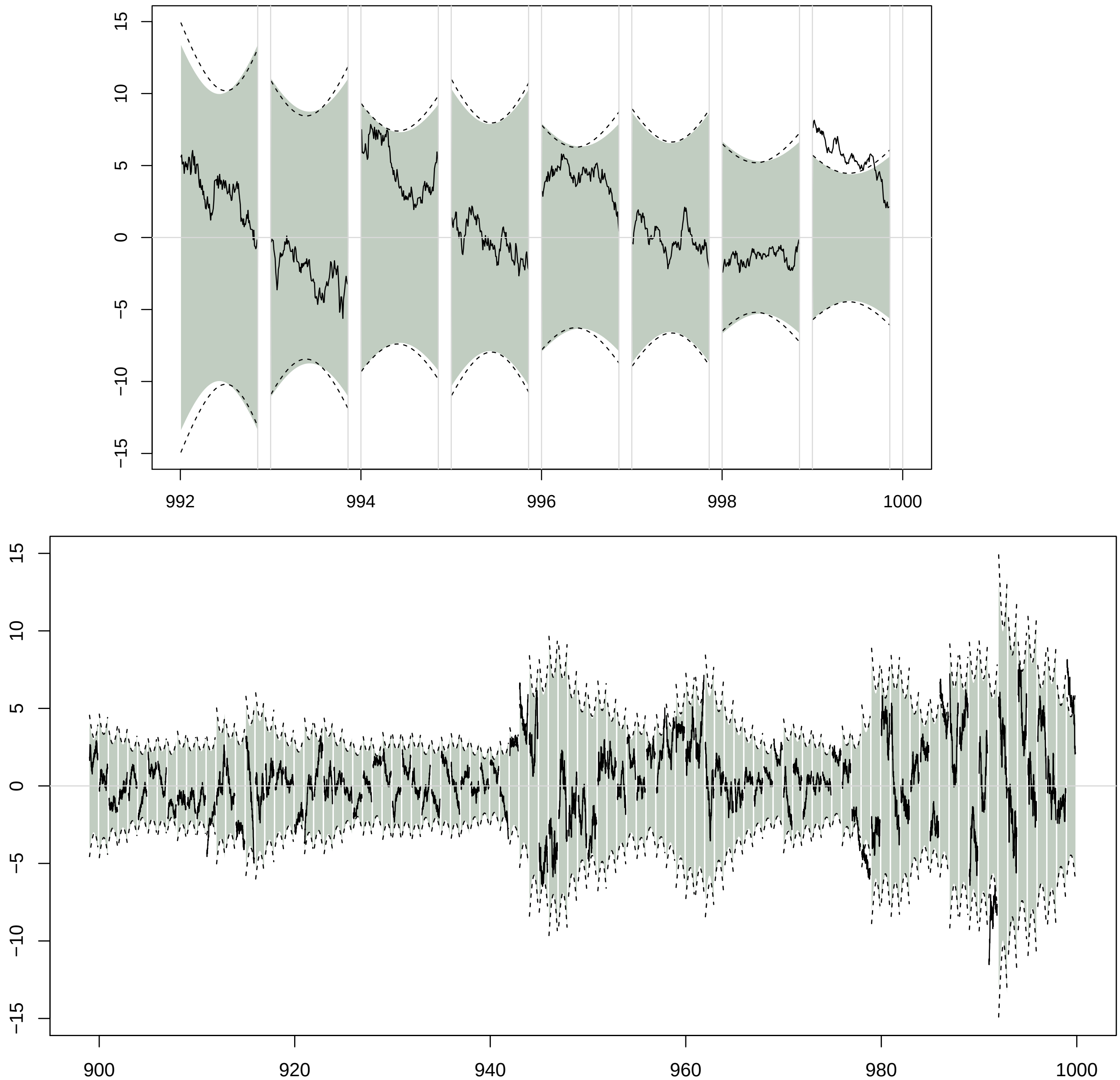

Figure 9: Solid lines represent the simulated process y_t, the shaded area is the region $\{[-2\sigma_t(u), 2\sigma_t(u)] \colon u \in [0,1]\}$. The dashed lines are estimators $\pm 2\widetilde{\sigma}_t(\hat{\theta})(u)$.

problem, we first observe that our model equations can be conveniently summarised in the following state-space form:

$$\underline{z}_t = \underline{b}_t + \mathbf{\Psi}_t(\underline{z}_{t-1}), \tag{4.4}$$

where $\underline{b}_t$, $\underline{z}_t \in H^{p+q}$ and $\mathbf{\Psi}_t \in \mathcal{L}(H^{p+q})$ are defined as

$$\underline{b}_t = (\eta_t^2 \delta, 0, \ldots, 0, \delta, 0, \ldots, 0)', \qquad \underline{z}_t = (y_t^2, \ldots, y_{t-q+1}^2, \sigma_t^2, \ldots, \sigma_{t-p+1}^2)',$$

and

$$\mathbf{\Psi}_t = \begin{pmatrix} \mathbf{\Upsilon}_t \boldsymbol{\alpha}_1 & \ldots & \mathbf{\Upsilon}_t \boldsymbol{\alpha}_{q-1} & \mathbf{\Upsilon}_t \boldsymbol{\alpha}_q & \mathbf{\Upsilon}_t \boldsymbol{\beta}_1 & \ldots & \mathbf{\Upsilon}_t \boldsymbol{\beta}_{p-1} & \mathbf{\Upsilon}_t \boldsymbol{\beta}_p \\ \boldsymbol{I}_H & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \cdots & \boldsymbol{I}_H & 0 & 0 & \cdots & 0 & 0 \\ \boldsymbol{\alpha}_1 & \ldots & \boldsymbol{\alpha}_{q-1} & \boldsymbol{\alpha}_q & \boldsymbol{\beta}_1 & \ldots & \boldsymbol{\beta}_{p-1} & \boldsymbol{\beta}_p \\ 0 & \cdots & 0 & 0 & \boldsymbol{I}_H & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & \boldsymbol{I}_H & 0 \end{pmatrix}.$$

Here, the operator $\mathbf{\Upsilon}_t$ is the pointwise multiplication by η_t^2, i.e. $H \ni x \longmapsto x\eta_t^2$. All 0's in the definition of the matrix $\mathbf{\Psi}_t$ are meant to be zero-operators.

Now, we introduce a mild technical assumption which we impose for the rest of the paper:

$$E \log^+ \|\eta_0^2\|_\infty < \infty, \tag{4.5}$$

where $\log^+(u) = \log(\max(1, u))$. By this assumption, it follows that $\|\eta_t^2\|_\infty < \infty$ a.s. and, hence, the linear operator $\mathbf{\Upsilon}_t$ is almost surely bounded. Indeed, we have that

$$\|\mathbf{\Upsilon}_t(x)\| = \|x\eta_t^2\| = \left(\int x^2(u)\eta_t^4(u)du\right)^{1/2} \leq \|\eta_t^2\|_\infty \|x\|, \quad \text{for any } x \in H,$$

thus

$$\|\mathbf{\Upsilon}_t\| \leq \|\eta_t^2\|_\infty. \tag{4.6}$$

For the sake of a light notation, we will now also use $\|\cdot\|$ for the norm on H^{p+q} as well as for the operator norm of $\mathcal{L}(H^{p+q})$. Its respective meaning will be clear from the context. From assumption (4.5) it is easily deduced that $E \log^+ \|\mathbf{\Psi}_1\| < \infty$. Moreover, the sequence $(\mathbf{\Psi}_t)$ is i.i.d. and our norm on $\mathcal{L}(H^{p+q})$ is sub-multiplicative. Hence, according to Theorem 6 in Kingman (1973) we almost surely have that

$$\gamma := \lim_{t\to\infty} \frac{1}{t} E(\log \|\mathbf{\Psi}_t \mathbf{\Psi}_{t-1} \cdots \mathbf{\Psi}_1\|) = \inf_{t\geq 1} \frac{1}{t} E(\log \|\mathbf{\Psi}_t \mathbf{\Psi}_{t-1} \cdots \mathbf{\Psi}_1\|) \tag{4.7}$$

$$= \lim_{t\to\infty} \frac{1}{t} \log \|\mathbf{\Psi}_t \mathbf{\Psi}_{t-1} \cdots \mathbf{\Psi}_1\|. \tag{4.8}$$

The coefficient $\gamma \in [-\infty, +\infty)$ is called the *top Lyapunov exponent* of the sequence $(\mathbf{\Psi}_t)_{t\in\mathbb{Z}}$.

Theorem 12. *Under* (4.5)*, a sufficient condition for the existence of a unique strictly stationary and non-anticipative solution to* (4.1)–(4.2) *is* $\gamma < 0$*.*

Proof. By iterating (4.4), we formally get that

$$\underline{z}_t = \underline{b}_t + \sum_{k=1}^{\infty} \mathbf{\Psi}_t \mathbf{\Psi}_{t-1} \cdots \mathbf{\Psi}_{t-k+1}(\underline{b}_{t-k}). \tag{4.9}$$

The series converges almost surely, since using (4.8), we deduce that

$$\limsup_{t\to\infty} \frac{1}{t} \log \|\mathbf{\Psi}_t \mathbf{\Psi}_{t-1} \cdots \mathbf{\Psi}_{t-k+1}(\underline{b}_{t-k})\| \leq \gamma + \limsup_{t\to\infty} \frac{1}{t} \log \|\underline{b}_{t-k}\|, \quad \text{a.s.} \tag{4.10}$$

Since $E \log^+ \|\underline{b}_{t-k}\| < \infty$ by (4.5), and $\|\underline{b}_{t-k}\| \geq \|\delta\| > 0$ the second summand is zero and, thus, we can apply the Cauchy rule to show convergence. In addition, it is easy to see that the $q+1$-th component of $\underline{z}_t$ defines a non-anticipative and stationary solution of (4.2). The proof of the existence is complete.

It remains to prove that the solution is almost surely unique. To this end, let us assume that $\underline{\tilde{z}}_t^*$ is another solution. By iterating (4.4), we get that

$$\underline{z}_t^* = \underline{z}_t^N + \mathbf{\Psi}_t \cdots \mathbf{\Psi}_{t-N}(\underline{z}_{t-N-1}^*), \text{ where } \underline{z}_t^N = \underline{b}_t + \sum_{k=1}^{N} \mathbf{\Psi}_t \cdots \mathbf{\Psi}_{t-k+1}(\underline{b}_{t-k}).$$

We then deduce that

$$\|\underline{z}_t^* - \underline{z}_t\| \leq \|\underline{z}_t^N - \underline{z}_t\| + \|\mathbf{\Psi}_t \mathbf{\Psi}_{t-1} \cdots \mathbf{\Psi}_{t-N}\| \cdot \|\underline{z}_{t-N-1}^*\|. \tag{4.11}$$

We already know that since $\gamma < 0$, we have $\|\underline{z}_t^N - \underline{z}_t\| \to 0$ and $\|\mathbf{\Psi}_t \cdots \mathbf{\Psi}_{t-N}\| \to 0$, almost surely when $N \to \infty$. Furthermore, the law of $\|\underline{z}_{t-N-1}^*\|$ is independent of N. Hence, the right-hand side of (4.11) tends to zero in probability. Therefore, $P(\underline{z}_t^* = \underline{z}_t) = 1$. □

Remark 4. Equations (4.7) and (4.8) are valid for any sub-multiplicative norm in $\mathcal{L}(H^{p+q})$, but the number γ will dependent on that choice (unless the norms are equivalent). For example we could define the top Lyapounov exponent γ_p associated to the following norm

$$\|\boldsymbol{\alpha}\|_{pp} = \sup_{\|x\|_{Lp}\leq 1} \|\boldsymbol{\alpha}(x)\|_{Lp}, \quad \forall\, p \in [1, +\infty].$$

Then we can show as in Theorem 12 that $\gamma_p < 0$ implies that a solution to (4.1)–(4.2) exists, but in $L^p[0,1]$. The case $p = 1$ might be interesting since it is more general than the case $p = 2$, because $L^2[0,1] \subset L^1[0,1]$. On the other hand, the case $p = \infty$ could be interesting as well since it provides the existence of a bounded solution (moreover the parameters might be chosen in such a way that the solution is continuous).

Remark 5. It would be interesting to see if the condition $\gamma < 0$ is also necessary for the existence of a strictly stationary solution to (4.1)–(4.2). The situation in the functional context is more complicated when compared to multivariate analysis. In the multivariate setup, one would argue that for some appropriately chosen matrix norm $\|\cdot\|_*$ we have that $\|\boldsymbol{\Psi}_t\boldsymbol{\Psi}_{t-1}\cdots\boldsymbol{\Psi}_{t-k+1}(\underline{b}_{t-k})\|_* \to 0$ (which is, of course, necessary for convergence of (4.9)) will imply $\|\boldsymbol{\Psi}_t\boldsymbol{\Psi}_{t-1}\cdots\boldsymbol{\Psi}_{t-k+1}\|_* \to 0$. In the infinite-dimensional setup, however, norms are not equivalent, and choosing a different norm will also give a different value for the exponent γ. In a second step, one uses contraction properties of random matrices in order to conclude. To extend such results to linear operators is beyond the scope of this paper.

In the next proposition, we specialise to the case of the functional GARCH(1,1) process in order to obtain a slightly more explicit result.

Proposition 4. *When $p = q = 1$, a sufficient condition for existence of a strictly stationary and non-anticipative solution to* (4.1)–(4.2) *is that*

$$E\log\left\|(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_{t-1}+\boldsymbol{\beta})\cdots(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_1+\boldsymbol{\beta})\right\| < 0, \quad \textit{for some } t \geq 1.$$

Proof. First, note that

$$\begin{aligned}\boldsymbol{\Psi}_t\boldsymbol{\Psi}_{t-1}\cdots\boldsymbol{\Psi}_1 &= \begin{bmatrix}\boldsymbol{\Upsilon}_t\\ \boldsymbol{I}_H\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha} & \boldsymbol{\beta}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Upsilon}_{t-1}\\ \boldsymbol{I}_H\end{bmatrix}\cdots\begin{bmatrix}\boldsymbol{\alpha} & \boldsymbol{\beta}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Upsilon}_1\\ \boldsymbol{I}_H\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha} & \boldsymbol{\beta}\end{bmatrix}\\ &= \begin{bmatrix}\boldsymbol{\Upsilon}_t\\ \boldsymbol{I}_H\end{bmatrix}(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_{t-1}+\boldsymbol{\beta})\cdots(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_1+\boldsymbol{\beta})\begin{bmatrix}\boldsymbol{\alpha} & \boldsymbol{\beta}\end{bmatrix},\end{aligned}$$

from which we can deduce a bound for the top Lyapunov exponent:

$$\begin{aligned}\gamma &\leq \lim_{t\to\infty}\frac{1}{t}\Big\{E\log(\left\|\eta_t^2\right\|_\infty+1)^{1/2} + E\log\left\|(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_{t-1}+\boldsymbol{\beta})\cdots(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_1+\boldsymbol{\beta})\right\|\\ &\qquad\qquad + E\log\left\|\begin{bmatrix}\boldsymbol{\alpha} & \boldsymbol{\beta}\end{bmatrix}\right\|\Big\}\\ &= \lim_{t\to\infty}\frac{1}{t}E\log\left\|(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_{t-1}+\boldsymbol{\beta})\cdots(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_1+\boldsymbol{\beta})\right\|\\ &= \inf_{t\geq 1}\frac{1}{t}E\log\left\|(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_t+\boldsymbol{\beta})\cdots(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_1+\boldsymbol{\beta})\right\|.\end{aligned}$$

The first inequality is in fact an equality since the two side terms are vanishing in the limit and the last equality follows from the Fekete's lemma. □

In their recent paper, Aue et al. (2016) obtained the condition

$$E\log\|\boldsymbol{\alpha}\boldsymbol{\Upsilon}_0+\boldsymbol{\beta}\|_{\mathcal{S}} < 0 \tag{4.12}$$

to guarantee a strictly stationary solution of functional GARCH(1,1) equations. Here, $\|\boldsymbol{\gamma}_0\|_{\mathcal{S}}$ is the Hilbert-Schmidt norm. Note that the Hilbert-Schmidt norm is dominating the operator norm and, hence,

$$\gamma \leq \frac{1}{t}E\log\left\|(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_t+\boldsymbol{\beta})\cdots(\boldsymbol{\alpha}\boldsymbol{\Upsilon}_1+\boldsymbol{\beta})\right\| \leq E\log\left\|\boldsymbol{\alpha}\boldsymbol{\Upsilon}_0+\boldsymbol{\beta}\right\| \leq E\log\|\boldsymbol{\alpha}\boldsymbol{\Upsilon}_0+\boldsymbol{\beta}\|_{\mathcal{S}}.$$

This shows that our condition is milder than that of Aue et al. (2016).

In the next proposition, we provide a sufficient condition for $E[y_t^2(u)] < \infty$ and equivalently $E[\sigma_t^2(u)] < \infty$, for all $u \in [0,1]$. We denote by $\rho(\boldsymbol{A})$ the spectral radius of the operator $\boldsymbol{A}$.

Proposition 5. *If $\gamma < 0$, then a sufficient condition for the existence of a pointwise second-order stationary solution to* (4.1)–(4.2) *is that $\rho(E\boldsymbol{\Psi}_0) < 1$.*

Proof. Since $\gamma < 0$, we deduce from Theorem 12 that there exists a strictly stationary and non-anticipative solution $(y_t)_{t\in\mathbb{Z}}$ to (4.1)–(4.2). Using stationarity and independence we deduce

$$E\underline{z}_t = E\underline{b}_t + \sum_{k=1}^{\infty} (E\boldsymbol{\Psi}_0)^k E\underline{b}_0, \tag{4.13}$$

which converges, since by assumption $\rho(E\boldsymbol{\Psi}_0) < 1$. This implies that $E[\sigma_t^2(u)] < \infty$ and $E[y_t^2(u)] < \infty$ for all $u \in [0,1]$. $\square$

We conclude this section with a result, which will be useful for statistical inference, but it also has its own interest.

Proposition 6. *Assume that $E\|\eta_0^2\|_\infty^\tau < \infty$ for some $\tau \in (0,1)$, $\gamma < 0$ and that $(y_t)_{t\in\mathbb{Z}}$ is a stationary solution to* (4.1)–(4.2)*. Then there exists $s \in (0,\tau)$ such that $E\|y_t^2\|^s < \infty$ and $E\|\sigma_t^2\|^s < \infty$.*

Proof. Using (4.7), there exists an integer t_0 such that $E\log\|\boldsymbol{\Psi}_{t_0}\boldsymbol{\Psi}_{t_0-1}\cdots\boldsymbol{\Psi}_1\|^\tau < 0$. Furthermore, we have that

$$E\|\boldsymbol{\Psi}_{t_0}\boldsymbol{\Psi}_{t_0-1}\cdots\boldsymbol{\Psi}_1\|^\tau \le E\|\boldsymbol{\Psi}_{t_0}\|^\tau\,\|\boldsymbol{\Psi}_{t_0-1}\|^\tau\cdots\|\boldsymbol{\Psi}_1\|^\tau = \left(E\|\boldsymbol{\Psi}_1\|^\tau\right)^{t_0},$$

where we used the fact that $(\boldsymbol{\Psi}_t)_{t\in\mathbb{Z}}$ are i.i.d. in the last equality (see Lemma 17 in the appendix for more detail on independence of random operators). Note that $E\|\boldsymbol{\Psi}_1\|^\tau < \infty$ by (4.6). From Lemma 2.2 in Francq and Zakoian (2011), we then deduce that there exists an $0 < s < \tau$ such that $\varsigma := E\|\boldsymbol{\Psi}_t\boldsymbol{\Psi}_{t-1}\cdots\boldsymbol{\Psi}_1\|^s < 1$. From (4.9) we get that

$$\begin{aligned} E\|\underline{z}_t\|^s &\le E\|\underline{b}_0\|^s\left\{1 + \sum_{k=1}^{\infty} E\|\boldsymbol{\Psi}_k\boldsymbol{\Psi}_{k-1}\cdots\boldsymbol{\Psi}_1\|^s\right\} \\ &\le E\|\underline{b}_0\|^s\left\{1 + \sum_{k=0}^{\infty}\varsigma^k\sum_{i=1}^{t_0}\left(E\|\boldsymbol{\Psi}_1\|^s\right)^i\right\}. \end{aligned}$$

Furthermore, we have that

$$E\|\underline{b}_0\|^s \le E\|\eta_0^2\delta\|^s + \|\delta\|^s \le \left(E\|\eta_0^2\|_\infty^s + 1\right)\|\delta\|^s < \infty.$$

Thus $E\|\underline{z}_t\|^s < \infty$ and the conclusion follows. $\square$

4.3 Estimation

A difficulty in estimating FDA models is that the concept of likelihood does not exist. For this reason, QML estimation cannot be straightforwardly defined in this framework. We propose an estimator which, though it cannot be related to any likelihood for the aforementioned reason, is directly inspired from the QMLE in the standard GARCH model. As regards the latter, the functional QMLE will be shown to be consistent in a semi-parametric framework in which the distribution of η_t does not need to be specified.

4.3.1 Parametrisation

From observations $(y_t)_{1\leq t\leq n}$ of curves satisfying Model (4.1)–(4.2), we consider inference on the parameters δ, $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_q$ and $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_p$. In order to guarantee identifiability of the model, we impose

$$E[\eta_0^2(u)] = 1,\ \forall u \in [0,1]. \tag{4.14}$$

An example of a stationary Gaussian process $(\eta_0(u))_{u\in[0,1]}$ satisfying (4.14) is the Ornstein-Uhlenbeck process given by $\eta_0(u) = e^{-u/2}W_0(e^u)$, where $W_0(\cdot)$ is the standard Brownian motion. This process has autocovariance function $\mathrm{Cov}(\eta_0(u+v),\eta_0(v)) = e^{-u/2}$. In general, however, we do not require either Gaussianity or “intraday–stationarity” of η_t.

We begin by assuming a specific parametrisation for Model (4.1)–(4.2). Let $\varphi_1,\dots,\varphi_M$ be linearly independent functions in H^+. We assume that there exists a non-negative valued vector $d = (d_1,\dots,d_M)'$ in $\mathbb{R}^M$, and non-negative valued matrices $A_i = (a_{k,\ell}^{(i)})$ and $B_j = (b_{k,\ell}^{(j)})$ in $\mathbb{R}^{M\times M}$ such that

$$\delta = \sum_{k=1}^{M} d_k\varphi_k, \quad \boldsymbol{\alpha}_i = \sum_{k,\ell=1}^{M} a_{k,\ell}^{(i)}\,\varphi_k\otimes\varphi_\ell \quad\text{and}\quad \boldsymbol{\beta}_j = \sum_{k,\ell=1}^{M} b_{k,\ell}^{(j)}\,\varphi_k\otimes\varphi_\ell, \tag{4.15}$$

for $i = 1,\dots q$, $j = 1,\dots p$. Note that $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_j$ belong to $\mathcal{K}_H^+$.[5] We define the parameter

$$\theta = \mathrm{vec}\big(d, A_1,\dots,A_q, B_1,\dots,B_p\big) \in \mathbb{R}^{M+(p+q)M^2}. \tag{4.16}$$

The model (4.1)–(4.2) is obtained for the value θ_0. By convention, we index by zero all quantities evaluated at θ_0. It is clear that the parametrisation is one-to-one in the sense that

$$\theta\neq\theta_0 \quad\Longrightarrow\quad (\delta,\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_p,\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_q) \neq (\delta_0,\boldsymbol{\alpha}_{01},\dots,\boldsymbol{\alpha}_{0p},\boldsymbol{\beta}_{01},\dots,\boldsymbol{\beta}_{0q}),$$

for $i = 1,\dots q$, $j = 1,\dots p$. To avoid confusion with the *parameter* θ we refer to δ, $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_q,\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_p$ as the *functional* parameters of the model. We assume that θ_0 belongs to a compact subset Θ of $\mathbb{R}_+^{M+(p+q)M^2}$, for a further discussion on the positivity of the functional parameters, see Section 5.3.3.

[5] Note that $K_{\boldsymbol{\alpha}_i}(u,v) = \sum_{k,\ell=1}^{M} a_{k,\ell}^{(i)}\varphi_k(u)\varphi_\ell(v)$ and $K_{\boldsymbol{\beta}_i}(u,v) = \sum_{k,\ell=1}^{M} b_{k,\ell}^{(i)}\varphi_k(u)\varphi_\ell(v)$.

Remark 6. The implication of equation (4.15) is that the volatility process $(\sigma_t^2)_{t\in\mathbb{Z}}$ belongs to the M-dimensional subspace of H spanned by $\varphi_1,\dots,\varphi_K$. This is also assumed in Aue et al. (2016). An alternative nonparametric approach will be developed in Section 4.4.

Our estimator is defined as follows:

$$\widehat{\theta}_n := \operatorname*{argmin}_{\theta\in\Theta} \widetilde{Q}_n(\theta), \tag{4.17}$$

where

$$\widetilde{Q}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \tilde{\ell}_t(\theta), \quad \tilde{\ell}_t(\theta) = \sum_{m=1}^{M}\left\{ \frac{\langle y_t^2, \varphi_m\rangle}{\langle \tilde{\sigma}_t^2, \varphi_m\rangle} + \log\langle \tilde{\sigma}_t^2, \varphi_m\rangle \right\}, \tag{4.18}$$

and where the empirical volatility $\tilde{\sigma}_t^2$ is computed recursively as

$$\tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\theta) = \delta + \sum_{i=1}^{q} \boldsymbol{\alpha}_i(y_{t-i}^2) + \sum_{j=1}^{p} \boldsymbol{\beta}_j(\tilde{\sigma}_{t-j}^2), \quad \text{for } t = 1,\dots,n \tag{4.19}$$

with some initial values $y_0,\dots,y_{-q+1}$ and $\tilde{\sigma}_0,\dots,\tilde{\sigma}_{-p+1}$ in H. Note that the positivity of the baseline functions φ_k ensures that the scalar products in (4.18) are positive and thus $\tilde{Q}_n(\theta)$ is well defined and reaches its minimum on the compact set Θ. This estimator is clearly inspired by the QMLE for standard GARCH models, and thus, we will refer to $\widehat{\theta}_n$ as the QML estimator.

4.3.2 Asymptotic results

Under (4.15), Model (4.1)-(4.2) admits a multivariate representation. More precisely, under the invertibility Assumption **A5** below, we define the process $(h_t(\theta))_{t\in\mathbb{Z}}$ as the stationary and ergodic solution of the following equation

$$h_t(\theta) = \mathfrak{d} + \sum_{i=1}^{q} \mathfrak{A}_i Y_{t-i}^{<2>} + \sum_{j=1}^{p} \mathfrak{B}_j h_{t-j}(\theta), \tag{4.20}$$

where $Y_t^{<2>} = (\langle y_t^2, \varphi_1\rangle, \dots \langle y_t^2, \varphi_M\rangle)'$, $\mathfrak{d} = \Phi d$ and for $i = 1,\dots,q$ and $j = 1,\dots,p$, $\mathfrak{A}_i = \Phi A_i$ and $\mathfrak{B}_j = \Phi B_j$ with $\Phi = ((\langle \varphi_i, \varphi_j\rangle))$ being the Gram-matrix of the functions $\varphi_1,\dots,\varphi_M$. Note that $h_t(\theta_0) = (\langle \sigma_t^2, \varphi_1\rangle, \dots, \langle \sigma_t^2, \varphi_M\rangle)'$.

We are able to deduce our main asymptotic results under the following assumptions:

A1 $\theta_0 \in \Theta$, Θ is a compact set.

A2 $E\|\eta_0^2\|_\infty^\tau < \infty$ for some $\tau \in (0,1)$, $(y_t)_{t\in\mathbb{Z}}$ is a strictly stationary and non-anticipative solution of Model (4.1)-(4.2).

A3 For any function $\psi \in H$ and any non-random constant κ,

$$\langle \eta_t^2, \psi\rangle = \kappa \ \text{ a.s.} \quad \Rightarrow \quad \psi \equiv 0 \ \text{ and } \ \kappa = 0.$$

A4 If $p > 0$, $\mathfrak{A}_0(z) = \sum_{i=1}^{q} \Phi A_{0i} z^i$ and $\mathfrak{B}_0(z) = I_M - \sum_{j=1}^{p} \Phi B_{0i} z^i$, are left co-primes and $[A_{0q}, B_{0p}]$ has full rank M.

A5 $\inf_{\theta\in\Theta} \mathfrak{d} > 0$ componentwise, and for all $\theta \in \Theta$, the matrix $\mathfrak{B}(z) = I_M - \sum_{j=1}^{p} \Phi B_i z^i$ is invertible for $|z| \leq 1$.

Theorem 13. *Under* (4.14)–(4.15) *and Assumptions* **A1-A5**, *the QMLE of* θ_0 *is strongly consistent, i.e. we have* $\widehat{\theta}_n \to \theta_0$ *almost surely.*

In order to derive the asymptotic law of our estimator we make the following assumptions:

A6 $\theta_0 \in \mathrm{Int}(\Theta)$.

A7 $E\|\eta_0\|_\infty^4 < \infty$.

Letting $\mathcal{N}_K(\mu, \Sigma)$ denote a K-variate normal random vector with mean μ and covariance matrix Σ, we get the following asymptotic normality result. Define $\ell_t(\theta)$ by replacing $\langle \tilde{\sigma}_t^2, \varphi_m \rangle$ by $h_{t,m}$ (the m-th component of $h_t(\theta)$) in $\tilde{\ell}_t(\theta)$.

Theorem 14. *Under* (4.14)–(4.15) *and Assumptions* **A1-A7** *we have that*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}_{M+(p+q)M^2}\left(0, J^{-1} I J^{-1}\right), \tag{4.21}$$

where $I = \mathrm{Var}\left(\frac{\partial \ell_t(\theta_0)}{\partial \theta}\right)$ *and*

$$J = E\left[\frac{\partial^2 \ell_t(\theta_0)}{\partial\theta\partial\theta'}\right] = \sum_{m=1}^{M} E\left[\frac{1}{h_{t,m}^2}\frac{\partial h_{t,m}}{\partial\theta}\frac{\partial h_{t,m}}{\partial\theta'}(\theta_0)\right]. \tag{4.22}$$

We remark that, unlike in the scalar case, it is not possible to factorise the matrix I in the asymptotic variance of Theorem 14. However, we have that

$$I = \sum_{m,m'=1}^{M} E\left[\left(\iint E[\eta_t^2(u)\eta_t^2(v)]\frac{\sigma_t^2(u)\sigma_t^2(v)}{h_{t,m}(\theta_0)h_{t,m'}(\theta_0)}\varphi_m(u)\varphi_{m'}(v)dudv - 1\right)\right.$$
$$\left. \times \frac{1}{h_{t,m}h_{t,m'}}\frac{\partial h_{t,m}}{\partial\theta}\frac{\partial h_{t,m'}}{\partial\theta'}(\theta_0)\right].$$

Using this, we can obtain estimates of J and I, as follows.

$$\widehat{J} = \frac{1}{n}\sum_{m=1}^{M}\sum_{t=1}^{n}\left[\frac{1}{\tilde{h}_{t,m}(\widehat{\theta}_n)\tilde{h}_{t,m'}(\widehat{\theta}_n)}\frac{\partial\tilde{h}_{t,m}(\widehat{\theta}_n)}{\partial\theta}\frac{\partial\tilde{h}_{t,m'}(\widehat{\theta}_n)}{\partial\theta'}\right]$$

and

$$\widehat{I} = \frac{1}{n}\sum_{m,m'=1}^{M}\sum_{t=1}^{n}\left[\left(\iint \widehat{K}_{\eta_0^2}(u,v)\frac{\tilde{\sigma}_t^2(\widehat{\theta}_n)(u)\,\tilde{\sigma}_t^2(\widehat{\theta}_n)(v)}{\tilde{h}_{t,m}(\widehat{\theta}_n)\tilde{h}_{t,m'}(\widehat{\theta}_n)}\varphi_m(u)\varphi_{m'}(v)dudv - 1\right)\right.$$
$$\left. \times \frac{1}{\tilde{h}_{t,m}(\widehat{\theta}_n)\tilde{h}_{t,m'}(\widehat{\theta}_n)}\frac{\partial\tilde{h}_{t,m}(\widehat{\theta}_n)}{\partial\theta}\frac{\partial\tilde{h}_{t,m'}(\widehat{\theta}_n)}{\partial\theta'}\right],$$

where the vector $\tilde{h}_t(\widehat{\theta}_n)$ and its derivatives are computed recursively by using equation (4.20),

$$\tilde{\sigma}_t^2(\widehat{\theta}_n)(u) = \sum_{m=1}^{M}(\Phi^{-1}\tilde{h}_t(\widehat{\theta}_n))_m\varphi_m(u),$$

and

$$\widehat{K}_{\eta_0^2}(u,v) = \frac{1}{n}\sum_{t=1}^{n} y_t^2(u)/\tilde{\sigma}_t^2(\widehat{\theta}_n)(u)\cdot y_t^2(v)/\tilde{\sigma}_t^2(\widehat{\theta}_n)(v).$$

4.3.3 Choice of instrumental functions

The instrumental functions $\varphi_1, \ldots, \varphi_M$ must satisfy the positivity constraints. We can consider any family of non-negative and linearly independent functions in H, such as the power basis $1, u, u^2, \ldots$ the exponential basis $e^u, e^{2u}, \ldots$ or some polynomial basis that is non-negative on $[0, 1]$, for example the popular B-splines bases. Since the latter are thought to perform well with functional data, we will consider them in our empirical study. More precisely, we will use the Bernstein polynomials. For more details on the use of B-splines and smoothing methods for functional data see Ramsay and Silverman (2006).

In order to avoid using a specific set of instrumental functions, we propose a heuristic data-driven method. Direct use of functional PCA (which is by far the most common practice in many applications) is not possible in this framework, due to the positivity constraints. Under the further assumption that $E\|y_t\|^4 < \infty$, we represent the squared process through its functional principal components, i.e.

$$y_t^2(u) = \mu(u) + \sum_{j=1}^{\infty} \langle y_t^2 - \mu, \psi_j \rangle \psi_j(u),$$

where $\mu(u) = E[y_t^2(u)]$ and $(\psi_j)_{j \geq 1}$ are the eigenfunctions of the covariance operator of y_t^2 (see, e.g. Horváth and Kokoszka (2012) for more details on functional principal components). Then, since $\sigma_t^2 = E[y_t^2 | \mathcal{F}_{t-1}]$ and in view of (4.2), it seems natural to assume that δ, $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_j$ are spanned by the finite set of functions $\mu, \psi_1, \ldots, \psi_{M-2}$. However, these functions are not non-negative. We, thus, propose to modify them according to the following routine. Take $\varphi_1(u) = 1$, $\varphi_2(u) = \mu(u)$, which is necessarily non-negative, and shift the other principal components, if necessary:

$$\varphi_m(u) := \psi_{m-2}(u) - \inf_{u \in [0,1]} \psi_{m-2}(u) \wedge 0, \text{ for all } \ m = 3, \ldots, M. \tag{4.23}$$

We have observed in our simulations (see Example 2) that this empirical choice performs relatively well, even when we compare it to the settings where the true (but unknown) basis functions in the data-generating process were used.

4.4 Extension to infinite-dimensional parameter space

Assuming a finite-dimensional parametrisation (4.15) may appear to be not entirely satisfactory from the theoretical standpoint. In this section, we show that the QML estimator remains strongly consistent in a more general setting, permitting an infinite-dimensional specification. For simplicity, we only consider the case when $p = q = 1$. We assume that δ, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be parametrised by some infinite-dimensional parameter $\theta \in \Theta$, in other words we let $M = \infty$ in (4.15). This parameter space is assumed to be a compact subset of l^2 (the set of square summable sequences).

Our new estimator is defined as

$$\widehat{\theta}_n^N := \operatorname*{argmin}_{\theta \in \Theta_N} \widetilde{Q}_n(\theta),$$

where $\Theta_N \subset \Theta$ is the subspace of all sequences with zero entries in components $k > N$. Furthermore, $\widetilde{Q}_n(\theta)$ is defined as in (4.18) and we set

$$\widetilde{\ell}_t(\theta) = \sum_{m=1}^{\infty} w_m \left\{ \frac{\langle y_t^2, \varphi_m \rangle}{\langle \tilde{\sigma}_t^2(\theta), \varphi_m \rangle} + \log \langle \tilde{\sigma}_t^2(\theta), \varphi_m \rangle \right\}, \tag{4.24}$$

where $(w_m)_{m\geq 1}$ is a non-negative and summable sequence of numerical weights.

Let $\boldsymbol{\alpha}^*$ denote the *adjoint* operator of $\boldsymbol{\alpha}$, i.e. the unique operator such that $\langle \boldsymbol{\alpha}^*(x), y\rangle = \langle x, \boldsymbol{\alpha}(y)\rangle$ for all $x, y \in H$. The following technical assumptions will be used.

A8 Identifiability. For all $m \geq 1$ and $\theta \in \Theta$ we have that

(a) $\langle \delta_0, \varphi_m\rangle = \langle \delta, \varphi_m\rangle$ implies $\delta = \delta_0$.
(b) $\boldsymbol{\alpha}_0^*(\varphi_m) = \boldsymbol{\alpha}^*(\varphi_m)$ implies $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$.
(c) $(\boldsymbol{\alpha}_0^* \circ \boldsymbol{\beta}_0^*)(\varphi_m) = (\boldsymbol{\alpha}_0^* \circ \boldsymbol{\beta}^*)(\varphi_m)$ implies $\boldsymbol{\beta} = \boldsymbol{\beta}_0$.

A9 There exists a sequence $(a_m)_{m\geq 1}$ such that w_m/a_m^2 is summable and $a_m \leq \int |\varphi_m(u)|du \leq \|\varphi_m\| \leq 1$ for all $m \geq 1$. Furthermore, for all $\theta \in \Theta$, the function δ_θ is uniformly bounded from below by some constant $c > 0$.

Remark 7. Note that, if $M < \infty$, $p = q = 1$, and $\mathfrak{A}_0$ is invertible then (4.15) implies **A8**. See Lemma 18 in the appendix for more details.

Proposition 7. *Under* (4.14) *and assumptions* **A1**$-$**A3***, and* **A8** $-$**A9***, the QMLE of* θ_0 *is strongly consistent:* $\widehat{\theta}_n^{N_n} \to \theta_0$ *almost surely in* l^2*, for any sequence* $N_n \nearrow \infty$*.*

To illustrate this result, we present an example of a functional GARCH process which is not included in the multivariate and linearly parametrised setting of Section 4.3.

Let $(\psi_k)_{k\geq 1}$ be an orthonormal basis of H. We assume that the volatility recursion is of the form

$$\sigma_t^2(u) = \exp\left(\sum_{k=1}^{\infty} d_k \psi_k(u)\right) + a\int y_{t-1}^2(v)dv + b\int \sigma_{t-1}^2(v)dv. \tag{4.25}$$

Here, the parameter is $\theta = (a, b, d_1, d_2, \dots) \in \mathbb{R}_+^2 \times \mathbb{R}^\infty$. The exponential function is used to guarantee a positive intercept, but other positive valued functions could be used instead of it. This model provides a very simple interpretation of the function δ. Indeed, we can show that the curve of δ parallels that of the expected intraday-volatility. More precisely, if $Ey_t^2(u) < \infty$, then we have that

$$E\sigma_t^2(u) = Ey_t^2(u) = \delta(u) + \frac{a+b}{1-a-b}\int \delta(v)dv.$$

We can also compute explicitly the top Lyapunov exponent which only depends on a, b and the law of η_0:

$$\begin{aligned}
\gamma &= \lim_{t\to\infty} \frac{1}{t} E\log \left\|(\boldsymbol{\alpha}\Upsilon_{t-1} + \boldsymbol{\beta})\cdots(\boldsymbol{\alpha}\Upsilon_1 + \boldsymbol{\beta})\right\| \\
&= \lim_{t\to\infty} \frac{1}{t} E\log \sup_{\|x\|\leq 1} \prod_{s=2}^{t-1} (a\int \eta_s^2(v)dv + b)\int (a\eta_1^2(v) + b)x(v)dv \\
&= E\log\left(a\int \eta_0^2(v)dv + b\right).
\end{aligned}$$

Finally, Proposition 7 can be applied to model (4.25). Indeed, we can easily choose a compact subset Θ of ℓ^2 and an innovation process $(\eta_t)_{t\in\mathbb{Z}}$ such that Assumptions **A1**$-$**A3**

	d	a	b
sd	2.4e-05	0.01	0.031
bias	1.6e-05	0.0041	0.011

Table 1: Performance of `LSE`.

	d	a	b
sd	7.7e-06	0.0051	0.0093
bias	3.1e-06	0.0017	0.0036

Table 2: Performance of `QMLE`.

are satisfied. Let $(\varphi_m)_{m\geq 1}$ be any family of non-negative functions spanning H. If we further assume that $(d_k)_{k\geq 1}$ is absolutely summable and that $\kappa = \sup_{k\geq 1} \|\psi_k\|_\infty < \infty$, then we get that $\delta(u) = \exp\left(\sum_{k=1}^\infty d_k \psi_k(u)\right) \geq \exp\left(-\kappa \sum_{k=1}^\infty |d_k|\right) > 0$. Now, since $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$ and $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, it easy to see that **A8** (b) is satisfied as well as **A8** (c), provided that $a_0 \neq 0$. Assumption **A8** (a) follows from the fact that $(\varphi_m)_{m\geq 1}$ is a basis of the space H. Then, for any weight sequence $(w_m)_{m\geq 1}$ which satisfies **A9** we get by Proposition 7 that the QMLE of model (4.25) is strongly consistent.

4.5 Empirical results

4.5.1 Simulations

We will first compare the Least Squares Estimator (`LSE`) of Aue et al. (2016) and our `QMLE` that is defined in (4.17). We will next compare the QML with given instrumental functions φ_m to the data-driven procedure described in Section 4.3.3 (we then refer to `QMLE*`).

Example 1.

The first setup is taken from Aue et al. (2016). They consider a GARCH(1,1) model with

$$\delta(u) = .01,\ K_{\boldsymbol{\alpha}}(u,v) = K_{\boldsymbol{\beta}}(u,v) = 12u(1-u)v(1-v), \tag{4.26}$$

for $u, v \in [0,1]$. For the innovations, Ornstein-Uhlenbeck processes are chosen. They are defined as $\eta_0(u) = e^{-u/2} W_0(e^u)$, where $(W_0(u))_{u\in[0,1]}$ is a Brownian motion.

The recursion starts at initial value $\sigma_0^2 := \delta$, and the first 1000 curves are discarded. Aue et al. (2016) project on one basis function $\varphi_1(u) = \sqrt{30}u(1-u)$, $u \in [0,1]$. It follows that $K_{\boldsymbol{\alpha}}(u,v) = a\,\varphi_1(u)\varphi_1(v)$, with $a = 0.4$ and $K_{\boldsymbol{\beta}}(u,v) = b\,\varphi_1(u)\varphi_1(v)$, with $b = 0.4$. Note that δ is not spanned by $\varphi_1(u)$ and that $d = \langle \delta, \varphi_1 \rangle \approx .009$. It is assumed that φ_1 is known and we estimate d, a and b. For the `LSE` we impose $|b| \leq .99$, whereas, for the `QMLE` we impose that $a \geq 0$ and $0 \leq b \leq .99$. In order to compare the performance of the two procedures, we consider 100 Monte-Carlo replications of our estimation experiment. The results of our simulations are displayed in Tables 1–2 and in Figure 10. We see that standard deviation and bias differ by a factor of 2 to 3 in favour of the `QMLE` methods.

Example 2.

We now illustrate our estimator in a slightly more complex example. We consider a functional GARCH(1,1) model with $\delta(u) = (u - .5)^2 + .1$,

$$K_{\boldsymbol{\alpha}}(u,v) = (u-.5)^2 + (v-.5)^2 + .2, \quad \text{and} \quad K_{\boldsymbol{\beta}}(u,v) = (u-.5)^2 + (v-.5)^2 + .4. \tag{4.27}$$

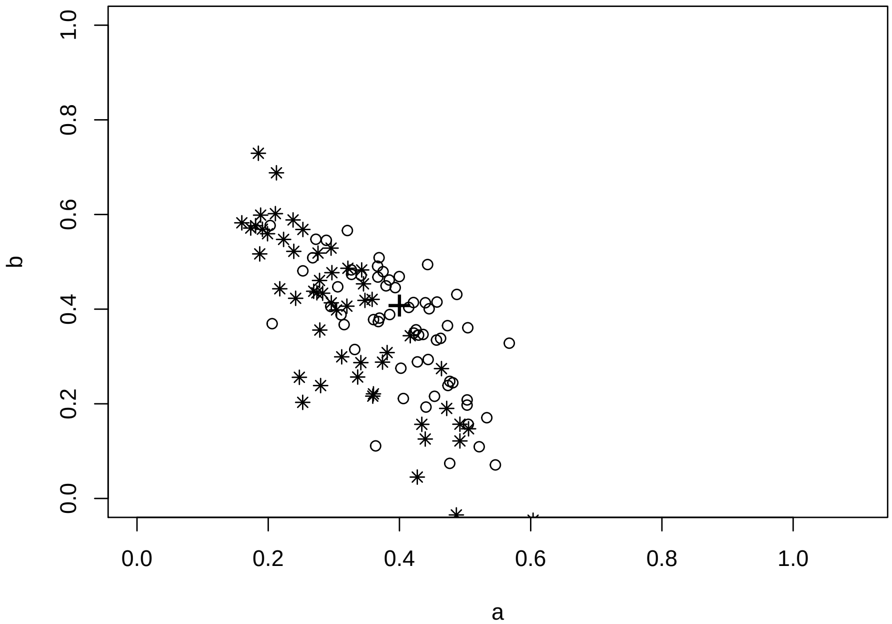

Figure 10: Estimates of a and b, with $*$ LSE and $\circ$ QMLE, and $+$ indicates the true values.

As in the previous example, we take for the innovations an i.i.d. sequence of Ornstein-Uhlenbeck processes, the recursion starts at initial value $\sigma_0^2 := \delta$, and the first 1000 curves are discarded.

For the instrumental functions $\varphi_1, \ldots, \varphi_M$ we consider the following families:

1. QMLE: Bernstein polynomials, which are a special case of B-spline functions defined by $\varphi_k(u) = \binom{M-1}{k-1} u^{k-1}(1-u)^{M-k}$, for $k = 1, \ldots, M$ and $u \in [0, 1]$.
2. QMLE*: The functions defined in Section 4.3.3.

We fix $M = 4$, then the subspace spanned by the Bernstein polynomials (of order 3) contains the true parameters defined in (4.27). We have constrained the parameters as follows: $d_k \geq 10^{-5}$, $a_{k\ell} \geq 0$ and to avoid an explosive solution, $0 \leq b_{k\ell} \leq (M \cdot \max_{1 \leq m \leq M} \|\varphi_m\|)^{-1}$, for all $k, \ell = 1, \ldots, M$.

We have represented the functional parameters δ and $\boldsymbol{\alpha}$ (its kernel $K_{\boldsymbol{\alpha}}$) in Figure 11 together with their QMLE. In order to compare the performance of the two procedures, we ran $N = 100$ Monte-Carlo replications of our estimation experiment with sample size $n = 1000$. The results of our simulations are displayed in Table 3. We show the relative mean squared deviations, defined as

$$\frac{1}{N^{1/2}\|\delta\|} \left(\sum_{\nu=1}^{N} \|\widehat{\delta}^{(\nu)} - \delta\|^2 \right)^{1/2}, \quad \frac{1}{N^{1/2}\|\boldsymbol{\alpha}\|} \left(\sum_{\nu=1}^{N} \|\widehat{\boldsymbol{\alpha}}^{(\nu)} - \boldsymbol{\alpha}\|^2 \right)^{1/2},$$

and analogously for $\boldsymbol{\beta}$. As aforementioned in Section 4.3.3 it is interesting that both procedures perform similarly, despite the fact that QMLE* doesn't require prior knowledge of instrumental functions.

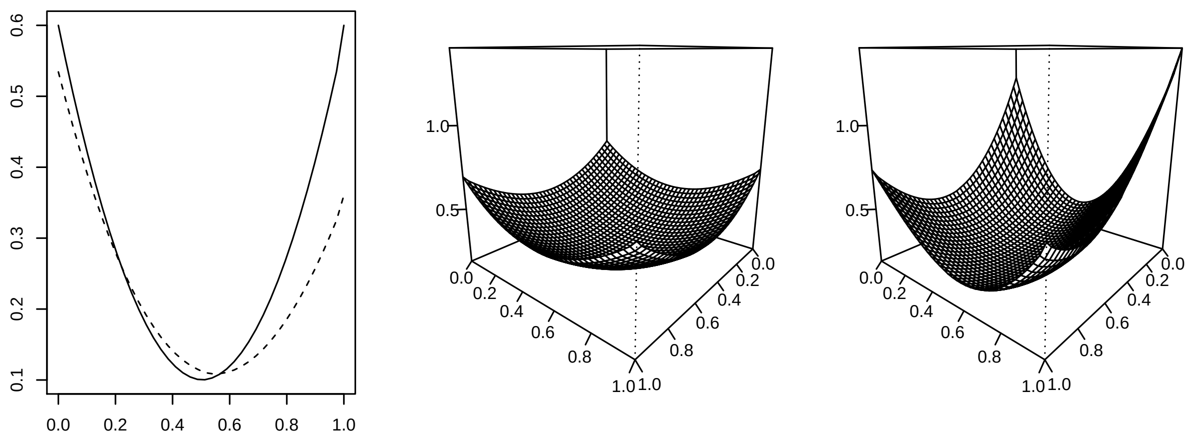

Figure 11: From left to right: the intercept function δ (solid line) compared to its estimation $\hat{\delta}$ (dashed line), the theoretical kernel $K_{\boldsymbol{\alpha}}(u, v)$ and the estimated kernel $K_{\widehat{\boldsymbol{\alpha}}}(u, v)$.

	δ	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$
`QMLE`	0.45	0.46	0.55
`QMLE*`	0.51	0.33	0.44

Table 3: Relative mean squared deviations for the corresponding functional parameters.

4.5.2 Real data illustration

We applied our estimators to the minutely recorded S&P100 Index for a ten-year period between 1997 and 2007. The return series is displayed in Figure 12. The functional

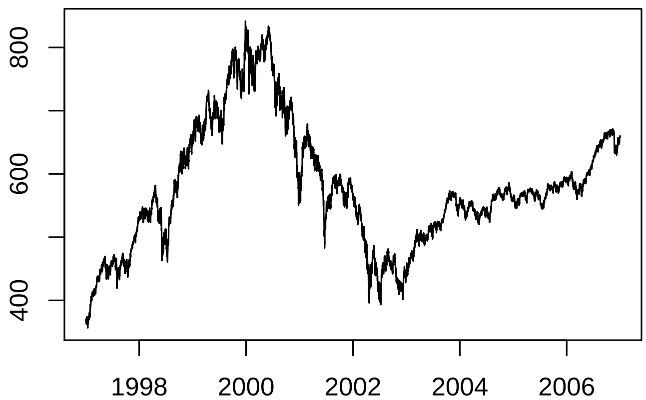

Figure 12: Raw data for S&P 100 index between 1997 and 2007.

GARCH model has been implemented on two different types of return data. Denoting by $X_t(u)$ the price at time u of the day t, we considered the τ-minute returns $y_t(u)$ and the intraday returns $\tilde{y}_t(u)$ defined by

$$y_t(u) = \log X_t(u) - \log X_t(u-\tau), \quad \text{and} \quad \tilde{y}_t(u) = \log X_t(u) - \log X_t(0). \tag{4.28}$$

For $y_t(u)$ we used $\tau = 20\,\text{min}$. For the instrumental functions we used, as in Example 2 the Bernstein polynomials with $M = 4$. We computed the `LSE` estimator to get an initial value of the parameter in the optimisation routine. The resulting empirical volatility curves are displayed in Figures 13 and 14 for y_t and in Figures 15 and 16 for $\tilde{y}_t$. In light of (4.3) and the related discussion, we plotted the curves $\tilde{\sigma}_t(\widehat{\theta}_n)(u) \cdot \widehat{Q}^{\hat{\eta}(u)}_{1-\alpha/2}$. The required quantiles were estimated from the residuals $\hat{\eta}_t(u) := y_t(u)/\tilde{\sigma}_t(\widehat{\theta}_n)(u)$, for $t = 1, \ldots, n$. On both processes we can observe the sensitivity to shocks of the volatility process and its persistence. The persistence seems stronger in Figure 16 than in Figure 14, whereas the rise of the volatility after a shock is more evident Figure 13 than in Figure 14. This is in line with a large value of $\|\widehat{\boldsymbol{\beta}}\|$ for $\tilde{y}_t$ (see Table 4).

	$\|\hat{\delta}\|$	$\|\hat{\boldsymbol{\alpha}}\|$	$\|\hat{\boldsymbol{\beta}}\|$
y	1e-06	0.46	0.46
$\tilde{y}$	5e-06	0.15	0.89

Table 4: Norms of the estimated parameters.

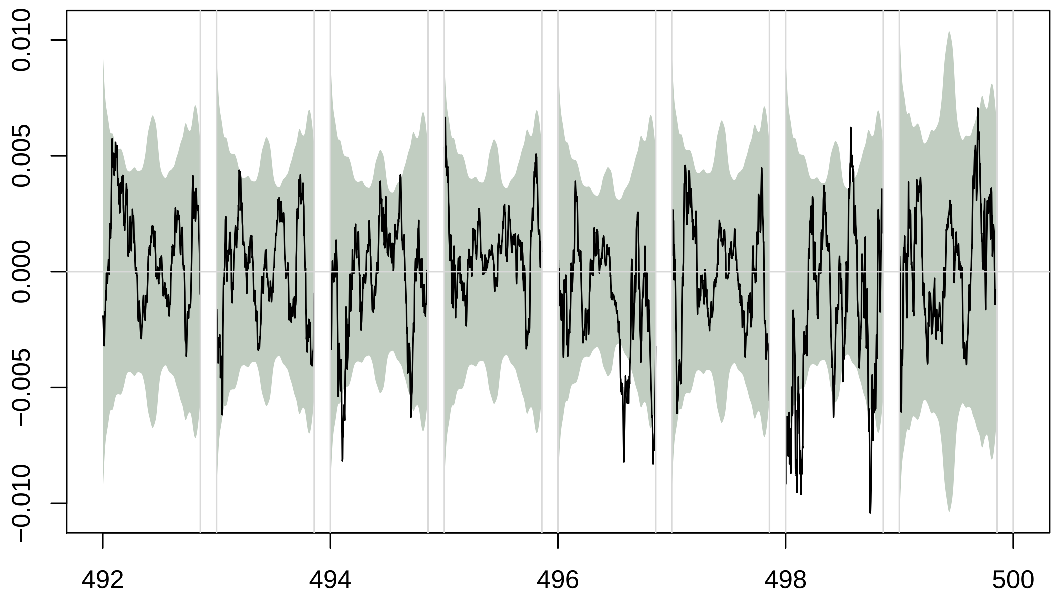

Figure 13: Predicted volatility (shaded area) for y_t (8 days)

Realised volatility

Practitioners often use the so-called *realised volatility* as a measure of the daily risk. Typically, it is defined as follows:

$$\mathrm{RV}_t = \sum_{j=1}^{\lfloor 1/\tau \rfloor} |\log X_t(j\tau) - \log X_t(j\tau - \tau)|^2. \tag{4.29}$$

If we choose the same τ as in the definition of y_t in (4.28) we remark that

$$\mathrm{RV}_t = \sum_{j=1}^{\lfloor 1/\tau \rfloor} |y_t^2(j\tau)| = \sum_{j=1}^{\lfloor 1/\tau \rfloor} \sigma_t^2(j\tau)\eta_t^2(j\tau).$$

At time $t-1$, the optimal predictor of RV_t is

$$E[\mathrm{RV}_t|\mathcal{F}_{t-1}] = \sum_{j=1}^{\lfloor 1/\tau \rfloor} \sigma_t^2(j\tau),$$

which can be estimated by

$$\widetilde{\mathrm{RV}}_t = \sum_{j=1}^{\lfloor 1/\tau \rfloor} \widetilde{\sigma}_t^2(\widehat{\theta})(j\tau), \tag{4.30}$$

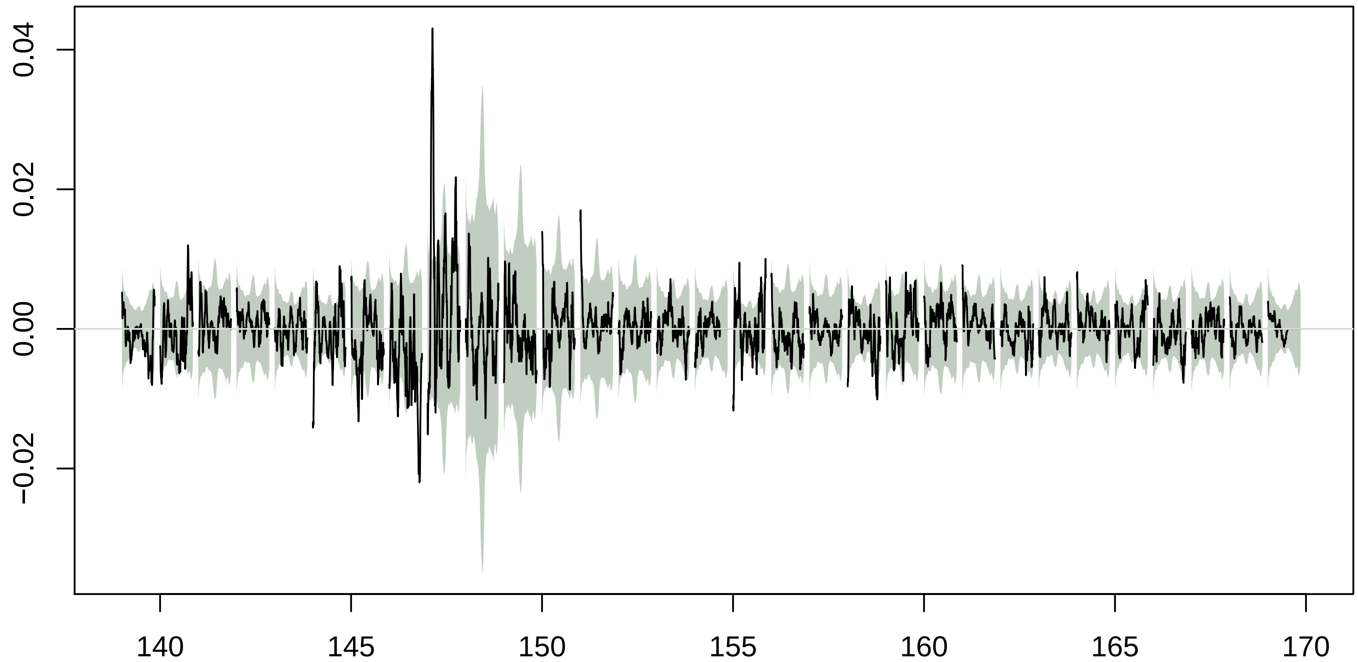

Figure 14: Predicted volatility (shaded area) for y_t (31 days).

where $\widehat{\theta}$ is the `QMLE` computed with the sub-sample $y_1, \ldots, y_{t-1}$. In Figure 17, we have plotted 41 one-day ahead predictions of $\widetilde{\mathrm{RV}}_t$ against RV_t.

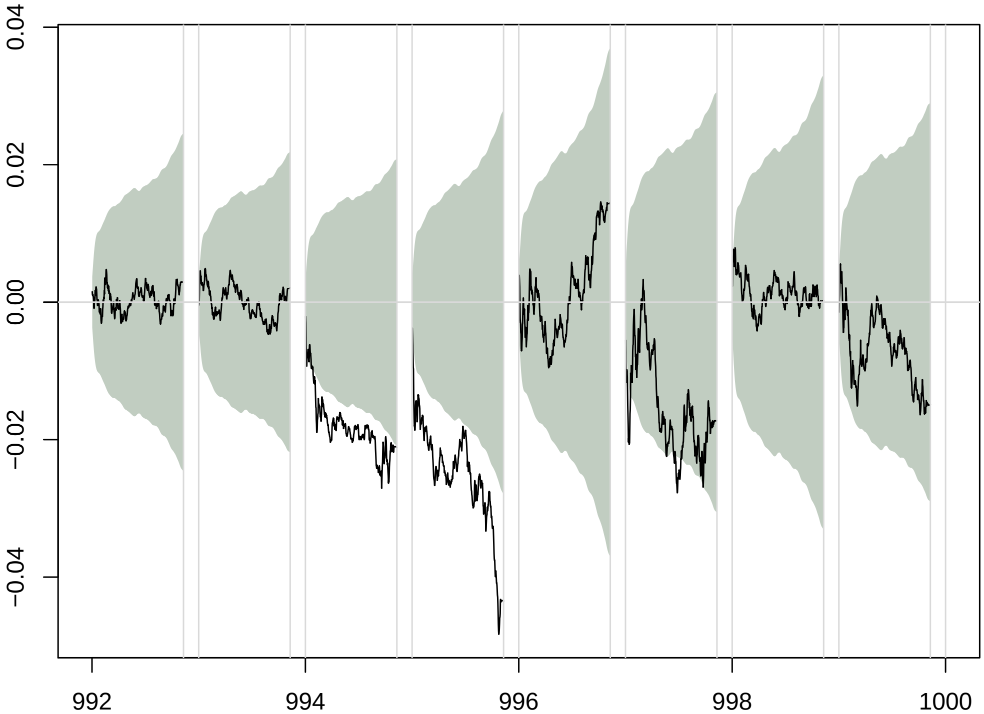

Figure 15: Predicted volatility (shaded area) for $\tilde{y}_t$ (8 days).

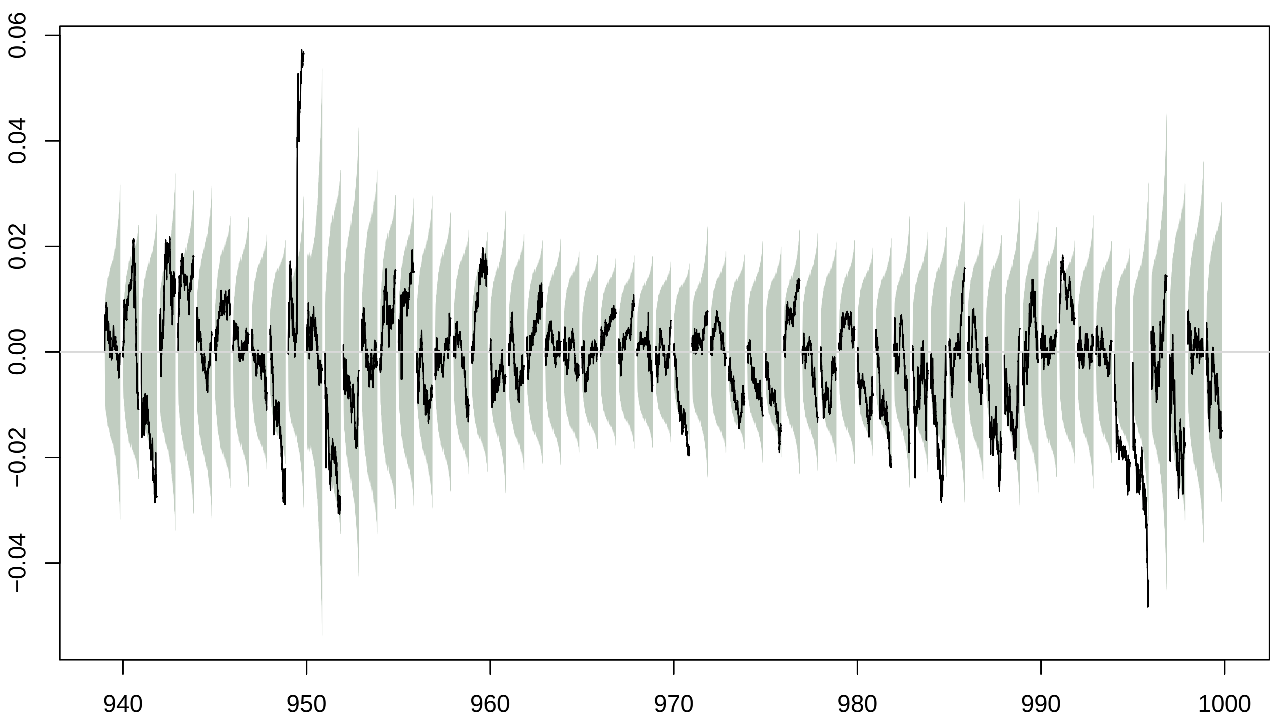

Figure 16: Predicted volatility (shaded area) for $\tilde{y}_t$ (61 days).

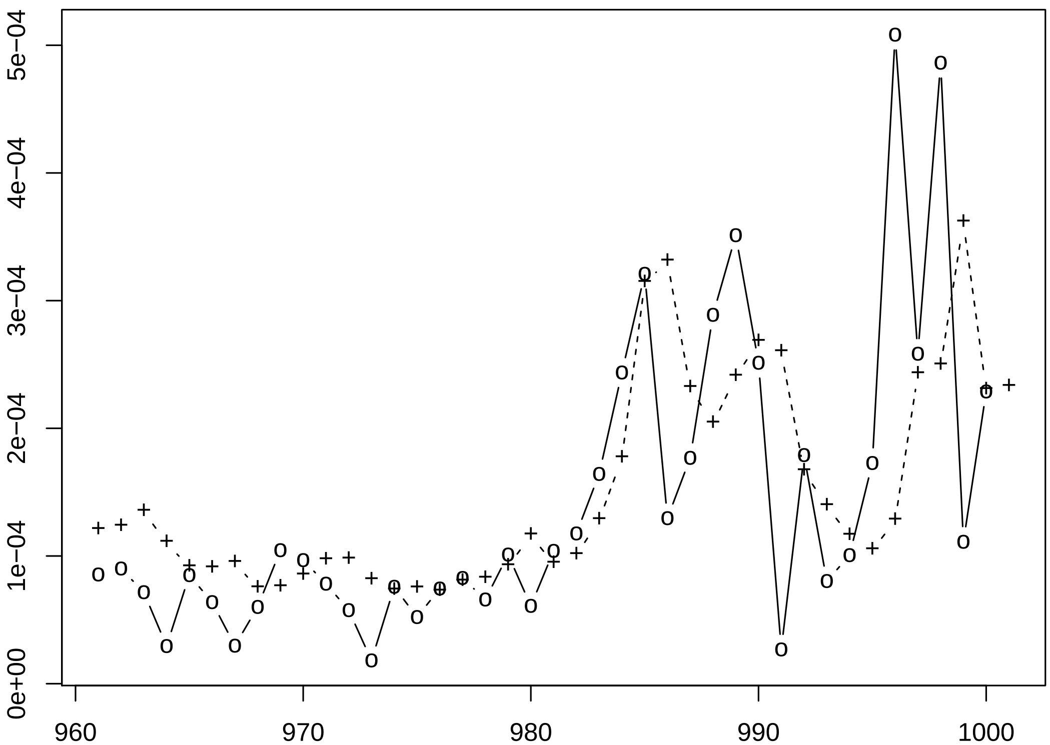

Figure 17: Predicted (+) and true realised volatility (∘).

4.6 Proofs of Chapter 4

We start to show asymptotic results for CCC GARCH models which will be used to prove Theorems 13 and 14. Similar results exist in the literature but with independent innovations. The convergence and asymptotic properties of semi-strong GARCH models has been considered by Escanciano (2009) in the scalar case. We provide a multivariate generalisation of this result to semi-strong CCC-GARCH models.

4.6.1 Asymptotics of semi-strong CCC-GARCH models

We recall the definition of a CCC-GARCH(p,q) process. It is an $\mathbb{R}^M$-valued process $(\epsilon_t)_{t\in\mathbb{Z}}$ with $\epsilon_t = (\epsilon_{t,1},\dots,\epsilon_{t,M})$ which satisfies the following equations:

$$\epsilon_t = H_t^{1/2}\nu_t, \tag{4.31}$$

$$H_t = D_t R D_t \quad \text{with} \quad D_t = \big(\operatorname{diag}(h_t)\big)^{1/2}, \tag{4.32}$$

$$h_t = \mathfrak{d} + \sum_{i=1}^{q} \mathfrak{A}_i \epsilon_{t-i}^{[2]} + \sum_{j=1}^{p} \mathfrak{B}_j h_{t-j}, \tag{4.33}$$

where $\epsilon_t^{[2]} = (\epsilon_{t,1}^2,\dots,\epsilon_{t,M}^2)$. Here R is an $M\times M$ correlation matrix, $\mathfrak{A}_i$ and $\mathfrak{B}_j$ are $M\times M$ matrices with positive elements, the components of the M-vector $\mathfrak{d}$ are strictly positive. We set

$$\xi = \big(\operatorname{vec}'(\mathfrak{d},\mathfrak{A}_1,\dots,\mathfrak{A}_q,\mathfrak{B}_1,\dots,\mathfrak{B}_p), r'\big)',$$

where r is the vector of the subdiagonal elements of R. The QMLE $\hat{\xi}_n$ of ξ_0 is defined by

$$\hat{\xi}_n = \arg\min_{\xi\in\Xi} \frac{1}{n}\sum_{t=1}^{n} \tilde{\ell}_t(\xi), \quad \tilde{\ell}_t(\xi) = \epsilon_t' \tilde{H}_t^{-1}\epsilon_t + \log|\det(\tilde{H}_t)|, \tag{4.34}$$

where $\tilde{H}_t$ is defined recursively using (4.33) and some initial values.

We define the matrix-valued polynomials $\mathfrak{A}(z) = \sum_{i=1}^{q}\mathfrak{A}_i z^i$ and $\mathfrak{B}(z) = I_M - \sum_{j=1}^{p}\mathfrak{B}_j z^j$, for $z\in\mathbb{C}$ and any ξ belonging to a compact parameter set Ξ. Let $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ be some filtration. The following technical assumptions are needed:

A*0 $E\|\epsilon_t^{[2]}\|^s < \infty$, for some $s>0$.

A*1 $\xi_0\in\Xi$, where Ξ is a compact set.

A*2 (ϵ_t) is a strictly stationary and ergodic solution of Model (4.31)-(4.33), with $\epsilon_t\in\mathcal{F}_t$.

A*3 $(\nu_t)_{t\in\mathbb{Z}}$ is an ergodic, stationary martingale difference sequence with respect to $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ such that $E[\nu_t\nu_t'|\mathcal{F}_{t-1}] = I_M$. There exists no vector $x\neq 0\in\mathbb{R}^M$ such that $x'\epsilon_t^{[2]}$ is $\mathcal{F}_{t-1}$-measurable.

A*4 If $q>0$, $\mathfrak{A}_0(z)$ and $\mathfrak{B}_0(z)$, are left co-primes and $[\mathfrak{A}_{0q},\mathfrak{B}_{0p}]$ has full rank M.

A*5 $\inf_{\xi\in\Xi}\mathfrak{d}>0$ componentwise; $\mathfrak{B}(z)$ is invertible for $|z|\leq 1$, for all $\xi\in\Xi$; R is a positive definite correlation matrix for all $\xi\in\Xi$.

A*6 $\xi_0\in\operatorname{Int}(\Xi)$.

A*7 $E\|\nu_t\nu_t'\|^{2(1+w)} < \infty$, for some $w > 0$.

Let $\ell_t(\xi) = \epsilon_t' H_t^{-1}(\xi)\epsilon_t + \log|\det(H_t(\xi))|$.

Theorem 15. *Under Assumptions* **A*0**–**A*5** *the QMLE of* ξ_0 *as defined in* (4.34) *is strongly consistent, i.e.* $\widehat{\xi}_n \to \xi_0 \quad \text{a.s.}$

Theorem 16. *Under Assumptions* **A*0**–**A*7**, *we have that*

$$\sqrt{n}(\hat{\xi}_n - \xi_0) \xrightarrow{d} \mathcal{N}_{M+(p+q)M^2+M(M-1)/2}\left(0, J^{-1}IJ^{-1}\right) \tag{4.35}$$

where $I = \operatorname{Var}\left(\frac{\partial \ell_t(\xi_0)}{\partial \xi}\right)$ *and* $J = E\left[\frac{\partial^2 \ell_t(\xi_0)}{\partial\xi\partial\xi'}\right]$.

Proof of Theorem 15. A close look into the proof of Theorem 11.7 in Francq and Zakoian (2011) shows that independence of the innovations is only needed to show the existence of some small order moments and for the identifiability. The existence of moments is now imposed in **A*0**. To show the identifiability, we have to prove that if there exists a matrix $P_1 \in \mathbb{R}^{M\times M}$ such that $P_1\epsilon_t^{[2]} = Z_{t-1}$, *a.s.* where the vector Z_{t-1} is $\mathcal{F}_{t-1}$ measurable, then $P_1 = 0$. Using the second part of our assumption **A*3** we may conclude. □

Proof of Theorem 16. As one can see in the proof of Theorem 11.8 in Francq and Zakoian (2011), the independence of the innovations is only used to show the existence of moments of the theoretical criterion ℓ_t, or of its derivative at ξ_0 or at a neighbourhood of it. For example, in order to prove the existence of I, we first compute for $i \le s := M + (p+q)M^2$,

$$\begin{aligned}
\frac{\partial \ell_t(\xi_0)}{\partial \xi_i} &= -\operatorname{Tr}\left\{(\epsilon_t\epsilon_t' D_t^{-1}R^{-1} + R^{-1}D_t^{-1}\epsilon_t\epsilon_t')D_t^{-1}\frac{\partial D_t(\xi_0)}{\partial \xi_i}D_t^{-1}\right\} + 2\operatorname{Tr}\left\{D_t^{-1}\frac{\partial D_t(\xi_0)}{\partial \xi_i}\right\} \\
&= \operatorname{Tr}\left\{(I_M - R^{-1/2}\nu_t\nu_t' R^{1/2})\frac{\partial D_t(\xi_0)}{\partial \xi_i}D_t^{-1} + (I_M - R^{1/2}\nu_t\nu_t' R^{-1/2})D_t^{-1}\frac{\partial D_t(\xi_0)}{\partial \xi_i}\right\}.
\end{aligned}$$

When independence between ν_t and the past holds, the existence of the second-order moments of these derivatives only requires $E\|\nu_t\|^4 < \infty$. Under our martingale difference assumption **A*3**, the moment condition on ν_t has to be strengthened as in **A*7**. More precisely, for $i, j \le s$, we use Hölder's inequality to get that

$$\begin{aligned}
E\left|\frac{\partial \ell_t(\xi_0)}{\partial \xi_i}\frac{\partial \ell_t(\xi_0)}{\partial \xi_j}\right| &\le \text{cst}\cdot\left(1 + E\|\nu_t\nu_t'\|^{2(1+w)}\right)^{\frac{1}{1+w}}\left(E\left\|D_t^{-1}\frac{\partial D_t(\xi_0)}{\partial \xi_i}\frac{\partial D_t(\xi_0)}{\partial \xi_j}D_t^{-1}\right\|^{\frac{1+w}{w}}\right)^{\frac{w}{1+w}} \\
&\le \text{cst}\cdot\left(E\left\|D_t^{-1}\frac{\partial D_t(\xi_0)}{\partial \xi_i}\right\|^{\frac{2(1+w)}{w}} E\left\|D_t^{-1}\frac{\partial D_t(\xi_0)}{\partial \xi_j}\right\|^{\frac{2(1+w)}{w}}\right)^{\frac{w}{2(1+w)}} < \infty.
\end{aligned}$$

The case $i > s$, i.e. when deriving with respect to the coefficients of the matrix R, is actually much simpler. The remainder of the proof works as in the classical CCC-GARCH by similar arguments. □

4.7 Proofs of the results of Section 4.3

In view of representation (4.20), we build a sequence (ϵ_t) satisfying the CCC-GARCH model of the previous section. Let $(r_t)_{t\in\mathbb{Z}}$ be an i.i.d. sequence of M-dimensional vectors, whose components are independent Rademacher variables. Let $\epsilon_t = \{\mathrm{diag}(Y_t^{<2>})\}^{1/2} r_t$ and let $\mathcal{F}_t$ the σ-field generated by $(r_t, \{\eta_{t-u}, u \geq 0\})$. Let $D_t = \{\mathrm{diag}(\langle \sigma_t^2, \varphi_m\rangle)\}^{1/2}$ and let $\nu_t = D_t^{-1}\epsilon_t$. Note that $\epsilon_t^{[2]} = (\epsilon_{t,1}^2, \ldots, \epsilon_{t,M}^2) = Y_t^{<2>}$. It follows that equations (4.31)-(4.33) hold with $R = I_M$. Let

$$\xi = \mathrm{vec}(\mathfrak{d}, \mathfrak{A}_1, \ldots, \mathfrak{A}_q, \mathfrak{B}_1, \ldots, \mathfrak{B}_p) = (I_{1+M(p+q)} \otimes \Phi)\theta,$$

where $\otimes$ denotes the usual Kronecker product of matrices. Since Φ is non-singular (this follows from the linear independence of the functions $\varphi_1, \ldots, \varphi_M$), the transformation T which maps θ to ξ is bijective. By choosing $\Xi = T(\Theta)$, the QML estimator $\hat{\xi}_n$ defined in (4.34) satisfies $\hat{\xi}_n = T(\hat{\theta}_n)$. Clearly, Theorems 15-16 can be straightforwardly adapted when $R = I$ is not estimated. It, therefore suffices, to verify that the assumptions of these theorems are satisfied.

Proof of Theorem 13. We start by verifying that the multivariate process $(\epsilon_t)_{t\in\mathbb{Z}}$ defined by $\epsilon_t = \{\mathrm{diag}(Y_t^{<2>})\}^{1/2} r_t$ satisfies assumptions $\mathbf{A}^*\mathbf{0}$–$\mathbf{A}^*\mathbf{5}$.

Assumption $\mathbf{A}^*\mathbf{0}$ follows from Proposition 6, noting that

$$\|\epsilon_t^{[2]}\| = \left(\sum_{m=1}^{M} |\langle y_t^2, \varphi_m\rangle|^2\right)^{1/2} \leq \|y_t^2\| \left(\sum_{m=1}^{M} \|\varphi_m\|^2\right)^{1/2},$$

where $\|.\|$ denotes the euclidean norm. Assumption $\mathbf{A}^*\mathbf{1}$ is obviously satisfied. By construction $\epsilon_t \in \mathcal{F}_t$ and satisfies (4.31)-(4.33). The stationarity and ergodicity of ϵ_t readily follows from $\mathbf{A2}$. Thus $\mathbf{A}^*\mathbf{2}$ holds. The first part of $\mathbf{A}^*\mathbf{3}$ is obtained by noting that $E(\langle y_t^2, \varphi_m\rangle^{1/2} r_{t,m}|\mathcal{F}_{t-1}) = 0$ and $E(\langle y_t^2, \varphi_m\rangle|\mathcal{F}_{t-1}) = \langle \sigma_t^2, \varphi_m\rangle$. For the second part of $\mathbf{A}^*\mathbf{3}$ we suppose that there exists an $x \in \mathbb{R}^M$, such that $x'\epsilon_t^{[2]}$ is $\mathcal{F}_{t-1}$-measurable. Then, conditionally on $\mathcal{F}_{t-1}$ we have that

$$x'\epsilon_t^{[2]} = \sum_{m=1}^{M} x_m \langle y_t^2, \varphi_m\rangle = \langle \eta_t^2, \sigma_t^2 \sum_{m=1}^{M} x_m\varphi_m\rangle = \text{const}, \quad \text{a.s.}$$

Assumption $\mathbf{A3}$ implies that the constant must be zero and that $\sigma_t^2(u)\sum_{m=1}^{M} x_m\varphi_m(u) = 0$, a.s. Furthermore, since $\sigma_t^2(u) \geq \delta_0(u) > 0$, for all $u \in [0,1]$ and the function $\varphi_1, \ldots, \varphi_M$ are linearly independent, we can conclude that $x = 0$. The remaining assumptions $\mathbf{A}^*\mathbf{4}$ and $\mathbf{A}^*\mathbf{5}$ are obviously satisfied.

Hence, we can apply Theorem 15 to the process $(\epsilon_t)_{t\in\mathbb{Z}}$ and, thus, we get that $\widehat{\xi}_n \to \xi_0$, a.s. To conclude, the continuity of T^{-1} implies that $\widehat{\theta}_n \to \theta_0$, a.s. □

Proof of Theorem 14. Since T is a bijection, it is obvious that assumption $\mathbf{A6}$ implies that $\mathbf{A}^*\mathbf{6}$. Next, we have

$$\begin{aligned}
E\|\nu_t\nu_t'\|^2 &\leq \sum_{k,m=1}^{M} E\nu_{t,k}^2\nu_{t,m}^2 \\
&\leq \sum_{k,m=1}^{M} \left(E\nu_{t,k}^4 E\nu_{t,m}^4\right)^{1/2} \\
&= \sum_{k,m=1}^{M} \left(E[E[\langle\sigma_t^2\eta_t^2,\varphi_k\rangle^2|\mathcal{F}_{t-1}]/\langle\sigma_t^2,\varphi_k\rangle^2]E[E[\langle\sigma_t^2\eta_t^2,\varphi_m\rangle^2|\mathcal{F}_{t-1}]/\langle\sigma_t^2,\varphi_m\rangle^2]\right)^{1/2} \\
&\leq M^2E\|\eta_t\|_\infty^4,
\end{aligned}$$

where we used Hölder's inequality in the last step. By assumption $\mathbf{A7}$ we obtain that $E\|\nu_t\nu_t'\|^2 < \infty$. This is slightly weaker than assumption $\mathbf{A}^*\mathbf{7}$, which would require more than fourth order moments for the innovations process. However, in our situation we can avoid this further assumption and the use of Hölder's inequality as in Theorem 16. For example, to prove that

$$E\left\|\frac{\partial\ell_t(\theta_0)}{\partial\theta}\frac{\partial\ell_t(\theta_0)}{\partial\theta'}\right\| < \infty \quad \text{and} \quad E\left\|\frac{\partial^2\ell_t(\theta_0)}{\partial\theta\partial\theta'}\right\| < \infty, \tag{4.36}$$

we compute

$$\frac{\partial\ell_t(\theta)}{\partial\theta} = \sum_{m=1}^{M}\left(1-\frac{Y_{t,m}^{<2>}}{h_{t,m}}\right)\frac{1}{h_{t,m}}\frac{\partial h_{t,m}}{\partial\theta}, \tag{4.37}$$

$$\begin{aligned}
\frac{\partial^2\ell_t(\theta)}{\partial\theta\partial\theta'} &= \sum_{m=1}^{M}\left(1-\frac{Y_{t,m}^{<2>}}{h_{t,m}}\right)\frac{1}{h_{t,m}}\frac{\partial^2 h_{t,m}}{\partial\theta\partial\theta'} \\
&\quad + \sum_{m=1}^{M}\left(2\frac{Y_{t,m}^{<2>}}{h_{t,m}}-1\right)\frac{1}{h_{t,m}^2}\frac{\partial h_{t,m}}{\partial\theta}\frac{\partial h_{t,m}}{\partial\theta'}.
\end{aligned} \tag{4.38}$$

Since, at the true value of the parameter $\theta = \theta_0$, we have that

$$\frac{Y_{t,m}^{<2>}}{h_{t,m_0}} = \frac{\langle y_t^2,\varphi_m\rangle}{\langle\sigma_t^2,\varphi_m\rangle} = \frac{\int\sigma_t^2(u)\eta_t^2(u)\varphi_m(u)du}{\int\sigma_t^2(u)\varphi_m(u)du} \leq \sup_{u\in[0,1]}\eta_t^2(u) = \|\eta_t^2\|_\infty. \tag{4.39}$$

This last quantity is independent of $\mathcal{F}_{t-1}$, which readily implies that (4.36) reduces to prove that

$$E\left|\frac{1}{h_{t,m}}\frac{\partial h_{t,m}}{\partial\theta_i}\right|^2 < \infty,$$

for all $m = 1,\ldots,M$ and $i = 1,\ldots,M+(p+q)M^2$. This can be established with Proposition 6. □

Proof of Proposition 7. We first prove that $\widehat{\theta}_n \to \theta_0$ almost surely, where $\widehat{\theta}_n$ denotes the minimiser of $\widehat{Q}_n$ over the whole space Θ. Let $\ell_t(\theta)$ denote the theoretical criterion (involving the infinite past of y_t), defined as $\tilde{\ell}_t(\theta)$, but with $\tilde{\sigma}_t^2(\theta)$ replaced by $\sigma_t^2(\theta)$ given by the model recursion. Note that under (4.14) we have that

$$E\left[\langle y_t^2,\varphi_m\rangle \mid \mathcal{F}_{t-1}\right] = \langle\sigma_t^2,\varphi_m\rangle. \tag{4.40}$$

To show $\widehat{\theta}_n \to \theta_0$ by standard arguments it suffices to verify the following:

(i) $\sup_{\theta\in\Theta}|Q_n(\theta)-\widetilde{Q}_n(\theta)| \underset{n\to\infty}{\longrightarrow} 0$ a.s.

(ii) $\langle\sigma_t^2(\theta),\varphi_m\rangle = \langle\sigma_t^2,\varphi_m\rangle$ a.s. for all $m\geq 1$ implies $\theta=\theta_0$;

(iii) $E\ell_t(\theta)$ exists for all $\theta\in\Theta$, and is finite for $\theta=\theta_0$, and $E\ell_t(\theta)>E\ell_t(\theta_0)$, for $\theta\neq\theta_0$;

(iv) $\forall\theta\neq\theta_0$, there exists a neighbourhood $\mathcal{V}_\theta$ such that $\liminf_n \inf_{\theta'\in\mathcal{V}_\theta}\widetilde{Q}_n(\theta') > E\ell_t(\theta_0)$, a.s.

Relation (iv) can be proven in the same way as in the univariate case. In order to prove (i), we recall that for all $\theta\in\Theta$, $\sigma_t^2(\theta)\geq\delta$. Hence, the non-negativity of the φ_m's and **A9** implies that there is a strictly positive constant c, such that $\langle\sigma_t^2(\theta),\varphi_m\rangle\geq ca_m$ for all $m\geq 1$. Furthermore, we have that $\|\sigma_t^2(\theta)-\tilde{\sigma}_t^2(\theta)\|\leq K\rho^t$, where ρ is the supremum of $\|\boldsymbol{\beta}_\theta\|$ over $\theta\in\Theta$ and K is a random constant. The compactness of Θ implies that $\rho<1$. We then compute

$$
\begin{aligned}
&\sup_{\theta\in\Theta}|Q_n(\theta)-\widetilde{Q}_n(\theta)|\\
&\leq \frac{1}{n}\sum_{t=1}^{n}\sum_{m=1}^{\infty} w_m \sup_{\theta\in\Theta}\left\{\left|\frac{\langle\sigma_t^2(\theta),\varphi_m\rangle-\langle\tilde{\sigma}_t^2(\theta),\varphi_m\rangle}{\langle\tilde{\sigma}_t^2(\theta),\varphi_m\rangle\langle\sigma_t^2(\theta),\varphi_m\rangle}\right|\langle y_t^2,\varphi_m\rangle - \log\frac{\langle\tilde{\sigma}(\theta)_t^2,\varphi_m\rangle}{\langle\sigma_t^2(\theta),\varphi_m\rangle}\right\}\\
&\leq \sum_{m=1}^{\infty} w_m K\left(\sup_{\theta\in\Theta}\frac{1}{\langle\delta_\theta,\varphi_m\rangle^2}\right)\frac{1}{n}\sum_{t=1}^{n}\rho^t\langle y_t^2,\varphi_m\rangle + \sum_{m=1}^{\infty} w_m K\left(\sup_{\theta\in\Theta}\frac{1}{\langle\delta_\theta,\varphi_m\rangle}\right)\frac{1}{n}\sum_{i=1}^{n}\rho^t\\
&\leq Kc^{-2}\sum_{m=1}^{\infty}\frac{w_m}{a_m^2}\frac{1}{n}\langle\sum_{t=1}^{n}\rho^t y_t^2,\varphi_m\rangle + Kc^{-1}\sum_{m=1}^{\infty}\frac{w_m}{a_m}\frac{1}{n}\sum_{t=1}^{n}\rho^t
\end{aligned}
$$

Consider the random function $Y=\lim_{n\to\infty}\uparrow\sum_{t=1}^{n}\rho^t y_t^2$. The existence of moment of order s for y_t^2 and its stationarity implies that $E\|Y\|^s\leq\sum_{t=1}^{\infty}\rho^{ts}E\|y_t^2\|^s<\infty$, and, thus, that Y is almost surely finite. We thus have that

$$
\sup_{\theta\in\Theta}|Q_n(\theta)-\widetilde{Q}_n(\theta)|\leq\frac{K\|Y\|}{nc^2}\sum_{m=1}^{\infty}\frac{w_m}{a_m^2}+\frac{K}{nc(1-\rho)}\sum_{m=1}^{\infty}\frac{w_m}{a_m}\ \underset{n\to\infty}{\overset{\text{a.s.}}{\longrightarrow}}\ 0.
$$

We now turn to (iii). Although $\ell_t(\theta)$ is not necessarily integrable, it is well defined in $\mathbb{R}\cup\{\infty\}$, since

$$
\begin{aligned}
E\ell_t(\theta)\geq E\sum_{m=1}^{\infty}w_m\log\langle\sigma_t^2(\theta),\varphi_m\rangle &\geq \sum_{m=1}^{\infty}w_m(\log c+\log a_m)\\
&\geq \text{const}-\sum_{m=1}^{\infty}w_m\log\Big(\frac{1}{a_m}\Big)>-\infty,
\end{aligned}
$$

and at the true value of the parameter $\theta=\theta_0$, we have that

$$
\begin{aligned}
E\,\ell_t(\theta_0)&=\sum_{m=1}^{\infty}w_m E\Big\{E\Big[\frac{\langle y_t^2,\varphi_m\rangle}{\langle\sigma_t^2,\varphi_m\rangle}|\mathcal{F}_{t-1}\Big]+\log\langle\sigma_t^2,\varphi_m\rangle\Big\}\\
&=\sum_{m=1}^{\infty}w_m\Big(1+E\log\langle\sigma_t^2,\varphi_m\rangle\Big)\leq\sum_{m=1}^{\infty}w_m\Big(1+E\log\|\sigma_t^2\|\Big)<\infty.
\end{aligned}
$$

We now have that

$$\begin{aligned} E[\ell_t(\theta)] - E[\ell_t(\theta_0)] &= \sum_{m=1}^{\infty} w_m E\left\{ \frac{\langle \sigma_t^2, \varphi_m\rangle}{\langle \sigma_t^2(\theta), \varphi_m\rangle} - 1 + \log \frac{\langle \sigma_t^2(\theta), \varphi_m\rangle}{\langle \sigma_t^2, \varphi_m\rangle} \right\} \\ &\geq 0, \end{aligned}$$

with equality if and only if for all $m \geq 1$

$$\langle \sigma_t^2(\theta), \varphi_m\rangle = \langle \sigma_t^2, \varphi_m\rangle, \quad \text{a.s.} \tag{4.41}$$

The proof of (iii) will be completed using (ii). To show (ii) we suppose that (4.40) holds true. We then have, for all $m \geq 1$,

$$\begin{aligned} &\langle \delta_0, \varphi_m\rangle + \langle \boldsymbol{\alpha}_0(y_{t-1}^2), \varphi_m\rangle + \langle \boldsymbol{\beta}_0(\sigma_{t-1}^2), \varphi_m\rangle \\ = \; &\langle \delta_\theta, \varphi_m\rangle + \langle \boldsymbol{\alpha}_\theta(y_{t-1}^2), \varphi_m\rangle + \langle \boldsymbol{\beta}_\theta(\sigma_{t-1}^2(\theta)), \varphi_m\rangle \;\text{ a.s.} \end{aligned} \tag{4.42}$$

We have,

$$\langle \boldsymbol{\alpha}_\theta(y_{t-1}^2), \varphi_m\rangle = \langle \sigma_{t-1}^2\eta_{t-1}^2, \boldsymbol{\alpha}_\theta^*(\varphi_m)\rangle = \langle \eta_{t-1}^2, \sigma_{t-1}^2 \boldsymbol{\alpha}_\theta^*(\varphi_m)\rangle.$$

In view of (4.42) we have

$$\langle \eta_{t-1}^2, \sigma_{t-1}^2[\boldsymbol{\alpha}_0^*(\varphi_m) - \boldsymbol{\alpha}_\theta^*(\varphi_m)]\rangle = K_{t-2} \quad \text{a.s.},$$

where $K_{t-2} \in \mathcal{F}_{t-2}$. It follows immediately that K_{t-2} must be constant and from **A3**, that

$$\boldsymbol{\alpha}_0^*(\varphi_m) = \boldsymbol{\alpha}_\theta^*(\varphi_m), \quad \forall m \geq 1.$$

By **A8** (b) we deduce that $\boldsymbol{\alpha}_\theta = \boldsymbol{\alpha}_0$. From Equation (4.42) we have, moreover, that

$$\langle \delta_0, \varphi_m\rangle + \langle \boldsymbol{\beta}_0(\sigma_{t-1}^2(\theta)), \varphi_m\rangle = \langle \delta_\theta, \varphi_m\rangle + \langle \boldsymbol{\beta}_\theta(\sigma_{t-1}^2), \varphi_m\rangle \;\text{ a.s.}$$

or, equivalently, that

$$\langle \delta_0, \varphi_m\rangle + \langle \sigma_{t-1}^2, \boldsymbol{\beta}_0^*(\varphi_m)\rangle = \langle \delta_\theta, \varphi_m\rangle + \langle \sigma_{t-1}^2(\theta), \boldsymbol{\beta}_\theta^*(\varphi_m)\rangle \;\text{ a.s.}$$

It follows that

$$\begin{aligned} &\langle \delta_0, \varphi_m\rangle + \langle \delta_{\theta_0} + \boldsymbol{\alpha}_0(y_{t-2}^2) + \boldsymbol{\beta}_{\theta_0}(\sigma_{t-2}^2), \boldsymbol{\beta}_0^*(\varphi_m)\rangle \\ &\qquad = \langle \delta_\theta, \varphi_m\rangle + \langle \delta_\theta + \boldsymbol{\alpha}_\theta(y_{t-2}^2) + \boldsymbol{\beta}_\theta(\sigma_{t-2}^2(\theta)), \boldsymbol{\beta}_\theta^*(\varphi_m)\rangle \;\text{ a.s.} \end{aligned}$$

Because $\boldsymbol{\alpha}_\theta = \boldsymbol{\alpha}_0$ we deduce, with obvious notation, that

$$\langle \boldsymbol{\alpha}_0(y_{t-2}^2), [\boldsymbol{\beta}_0^*(\varphi_m) - \boldsymbol{\beta}_\theta^*(\varphi_m)]\rangle = K_{t-3}, \;\text{ a.s.}$$

or, equivalently, that

$$\langle \eta_{t-2}^2, \sigma_{t-2}^2(\boldsymbol{\alpha}_0^* \circ \boldsymbol{\beta}_0^*)(\varphi_m)\rangle = \langle \eta_{t-2}^2, \sigma_{t-2}^2(\boldsymbol{\alpha}_0^* \circ \boldsymbol{\beta}_\theta^*)(\varphi_m)\rangle + K_{t-3}, \;\text{ a.s.}$$

With similar arguments as in the above we, thus, obtain $(\boldsymbol{\alpha}_0^* \circ \boldsymbol{\beta}_0^*)(\varphi_m) = (\boldsymbol{\alpha}_0^* \circ \boldsymbol{\beta})(\varphi_m)$, for all $m \geq 1$. By **A8** (c), this entails $\boldsymbol{\beta}_\theta = \boldsymbol{\beta}_0$. Finally, equation (4.42) reduces to

$\langle \delta_0, \varphi_m \rangle = \langle \delta_\theta, \varphi_m \rangle$, which, by **A8** (a), implies that $\delta = \delta_0$. We can conclude that $\theta = \theta_0$.

Recall that $\widehat{\theta}_n$ denotes the minimiser of $\widehat{Q}_n$ over the whole space Θ which is not computable in practise due to our assumption that Θ is infinite dimensional. Let us denote $\|x\|$ the ℓ^2-norm of some square summable sequence $x = (x_1, x_2, \ldots)$. We have that

$$\|\widehat{\theta}_n^{N_n} - \theta_0\| \leq \|\widehat{\theta}_n^{N_n} - \widehat{\theta}_n|_{\Theta_{N_n}}\| + \|\widehat{\theta}_n|_{\Theta_{N_n}} - \widehat{\theta}_n\| + \|\widehat{\theta}_n - \theta_0\|, \tag{4.43}$$

where $|_{\Theta_N}$ denotes the projection $x \mapsto (x_1, \ldots, x_N)$. Up to now, we showed that the third term of (4.43) converges almost surely to zero. The second term is equal to $\sum_{j>N_n} \widehat{\theta}_{n,j}$ where $\widehat{\theta}_{n,j}$ simply denotes the j-th term of the sequence $\widehat{\theta}_n$ which is supposed to be in ℓ^2. We can further bound this quantity by $\sup_{\ell \geq 1} \sum_{j>N_n} \widehat{\theta}_{\ell,j}$. To show that this converges to zero we apply the tighness Lemma 14 in Cerovecki and Hörmann (2017) with $p_j^{(n)} = \widehat{\theta}_{n,j}$ and $p_j^{(0)} = \theta_{0,j}$. Finally, from the compactness of Θ we know that there exists a subsequence $(\widehat{\theta}_{n_\ell}^{N_{n_\ell}})_{\ell \geq 1}$ that converges in ℓ^2 to x, say, and observe that by definition

$$\widetilde{Q}_{n_\ell}(\widehat{\theta}_{n_\ell}^{N_{n_\ell}}) \leq \widetilde{Q}_{n_\ell}(\widehat{\theta}_{n_\ell}|_{\Theta_{N_{n_\ell}}}), \quad \text{for all } \ \ell \geq 1.$$

Now, we have already shown that $\widehat{\theta}_{n_\ell}|_{\Theta_{N_{n_\ell}}} \to \theta_0$ a.s. when $\ell \to \infty$. Since $\widetilde{Q}_n(\theta) \to Q(\theta)$ a.s. and uniformly in θ, we obtain that $Q(x) \leq Q(\theta_0)$, a.s. This shows that $x = \theta_0$ and, thus, that the first term on the right-hand side of (4.43) converges a.s. to zero. $\square$

5 Appendix

5.1 Random elements in complex Hilbert spaces

In this section we present a framework for studying random elements in complex Hilbert spaces, which is analogue to the complex multivariate one, see e.g. Brockwell and Davis (1991) or Picinbono (1996). We assume that H is a separable complex Hilbert space and that there exists a separable real Hilbert space H_0, such that H is isomorphic as a vector space to $H_0 \oplus \mathrm{i}H_0$, i.e. every element $u \in H$ as a unique representation $u = u_1 + u_2$, where $u_1 \in H_0$ and $u_2 \in \mathrm{i}H_0$. In this framework, one can define explicitly the complex conjugate as $\overline{u} := u_1 - u_2$. A random element in H is a measurable mapping $Z : (\Omega, \mathcal{A}, P) \longrightarrow (H, \mathcal{B}_H)$. We usually denote its real and imaginary part by $X = \operatorname{Re} Z$ and $Y = \operatorname{Im} Z$. When there is no further specification $\langle \cdot, \cdot \rangle$ denotes the inner product of H.

The variance structure of a complex random variable U, i.e. a random element in $\mathbb{C}$, can be summarized by its *complex variance* $\gamma = E|U|^2$ together with its *relation* $c = EU^2$. This is just a complex formulation of the variance matrix from the bivariate random vector $(\mathrm{Re}(U), \mathrm{Im}(U))'$, the which can be obtain by

$$\Sigma = \frac{1}{2} \begin{pmatrix} \gamma + \mathrm{Re}(c) & \mathrm{Im}(c) \\ \mathrm{Im}(c) & \gamma - \mathrm{Re}(c) \end{pmatrix}. \tag{5.1}$$

In the functional setting, we define $\Gamma = EZ \otimes Z$ and $C = EZ \otimes \overline{Z}$. One can easily obtain that

$$\begin{aligned} \Gamma &= EX \otimes X + EY \otimes Y + \mathrm{i}(EY \otimes X - EX \otimes Y), \\ C &= EX \otimes X - EY \otimes Y + \mathrm{i}(EY \otimes X + EX \otimes Y). \end{aligned}$$

We define $\operatorname{Re} \Gamma := EX \otimes X + EY \otimes Y$, $\operatorname{Im} \Gamma := EY \otimes X - \mathbb{E}X \otimes Y$, $\operatorname{Re} C := EX \otimes X - \mathbb{E}Y \otimes Y$, and $\operatorname{Im} C := EY \otimes X + \mathbb{E}X \otimes Y$. We say that Z is a Gaussian random variable in the complex Hilbert space H if $\operatorname{Re} Z$ and $\operatorname{Im} Z$ are jointly Gaussian in $H_0 \times H_0$ and write $Z \sim \mathcal{CN}_H(0, \Gamma, C)$, where Γ and C are defined as above. The next proposition gives different characterization of Gaussianity in H and ensures that Γ and C indeed determine the distribution of Z.

Proposition 8. *Let Z be a random element in H with zero mean. The following assertions are equivalents.*

(i) $(\operatorname{Re} Z_n, \operatorname{Im} Z_n)' \sim \mathcal{N}_{H_0 \times H_0}(0, \Sigma)$, *for some* $\Sigma \in \mathcal{L}_{H_0 \times H_0}$.

(ii) *For every* $w \in H_0 \times H_0$, $\langle w, (\operatorname{Re} Z, \operatorname{Im} Z) \rangle_{H_0 \times H_0} \sim \mathcal{N}_{\mathbb{R}}(0, \sigma_w^2)$, *for some* $\sigma_w^2 > 0$.

(iii) $Z \sim \mathcal{CN}_H(0, \Gamma, C)$, *for some* Γ *and* $C \in \mathcal{L}_H$.

(iv) *For every* $u \in H$, $\langle Z, u \rangle \sim \mathcal{CN}_{\mathbb{C}}(0, \gamma_u, c_u)$ *for some* $\gamma_u > 0$ *and* $c_u \in \mathbb{C}$.

The corresponding result when $\mu = EZ \neq 0$, can be deduced straightforwardly. Furthermore, we have that $\sigma_w^2 = \langle \Sigma w, w \rangle_{H_0 \times H_0}$, $\gamma_u = \langle \Gamma u, u \rangle$, $c_u = \langle C \overline{u}, u \rangle$, and

$$\Sigma = \frac{1}{2} \begin{pmatrix} \operatorname{Re} \Gamma + \operatorname{Re} C & -\operatorname{Im} \Gamma + \operatorname{Im} C \\ \operatorname{Im} \Gamma + \operatorname{Im} C & \operatorname{Re} \Gamma - \operatorname{Re} C \end{pmatrix}.$$

Proof. It is quite clear that *(ii)* $\iff$ *(i)* $\stackrel{\text{def.}}{\iff}$ *(iii)* $\implies$ *(iv)*. We thus assume that *(iv)* holds. In other words, we have that

$$\begin{pmatrix} \mathrm{Re}\langle Z,u\rangle \\ \mathrm{Im}\langle Z,u\rangle \end{pmatrix} \sim \mathcal{N}_{\mathbb{R}^2}\left(0\,,\ \frac{1}{2}\begin{pmatrix} \gamma_u + \mathrm{Re}(c_u) & \mathrm{Im}(c_u) \\ \mathrm{Im}(c_u) & \gamma_u - \mathrm{Re}(c_u) \end{pmatrix}\right).$$

Therefore any linear combination of this is a univariate Gaussian variable. We define $u = u_0 + \mathrm{i}u_1$, then for all $s,t \in \mathbb{R}$ we have

$$s\mathrm{Re}\langle Z,u\rangle + t\mathrm{Im}\langle Z,u\rangle = [su_0 - tu_1, su_1 + tu_0]\begin{pmatrix} \mathrm{Re}\, Z \\ \mathrm{Im}\, Z \end{pmatrix}.$$

It is now very clear that $\mathrm{Re}(Z)$ and $\mathrm{Im}(Z)$ are jointly Gaussian, and thus $Z \sim \mathcal{CN}_H(0,\Gamma,C)$, where $\Gamma = EZ\otimes Z$ and $C = EZ\otimes\overline{Z}$. To conclude, we check that

$$\gamma_u = E|\langle Z,u\rangle|^2 = E\langle Z,u\rangle\langle u,Z\rangle = E\langle Z\langle u,Z\rangle, u\rangle = \langle \Gamma u,u\rangle,$$

and similarly

$$c_u = E\langle Z,u\rangle^2 = E\langle Z\langle Z,u\rangle,u\rangle = E\langle Z\langle \overline{u},\overline{Z}\rangle,u\rangle = \langle C\overline{u},u\rangle.$$

Moreover, to see that Γ and C are well determined by so to say, their diagonal values $\langle \Gamma u,u\rangle$ and $\langle C\overline{u},u\rangle$, one can use the following polarization identities

(5.2)
$$\langle \Gamma u,v\rangle = \frac{1}{4}\Big[\langle \Gamma(u+v),u+v\rangle - \langle \Gamma(u-v),u-v\rangle + \mathrm{i}\langle \Gamma(u+\mathrm{i}v),u+\mathrm{i}v\rangle - \mathrm{i}\langle \Gamma(u-\mathrm{i}v),u-\mathrm{i}v\rangle\Big]$$
$$\langle C\overline{u},v\rangle = \frac{1}{4}\Big[\langle C(\overline{u+v}),u+v\rangle - \langle C(\overline{u-v}),u-v\rangle + \mathrm{i}\langle C(\overline{u+\mathrm{i}v}),u+\mathrm{i}v\rangle - \mathrm{i}\langle C(\overline{u-\mathrm{i}v}),u-\mathrm{i}v\rangle\Big],$$

for all $u,v \in H$. Note that the variance Γ is self adjoint operator i.e. $\langle \Gamma u,v\rangle = \langle u,\Gamma v\rangle$, whereas the relation operator C satisfies $\langle Cu,v\rangle = \langle u,\overline{C\overline{v}}\rangle$. □

Similarly, we can show that Z_1 and Z_2 are *non correlated* if and only if both $EZ_1\otimes Z_2$ and $EZ_1\otimes\overline{Z_2}$ are the zero operator. We deduce the following complex version of the CLT for *martingale difference sequences* (MDS).

Theorem 17. *Let $(Z_t)_{t\geq 1}$ be complex valued MDS with $E|Z_1|^2 < \infty$. We further assume that there exists constants $\gamma > 0$ and $c \in \mathbb{C}$, such that*

$$\frac{1}{n}\sum_{t=1}^{n} E\left[|Z_t|^2|\mathcal{F}_{t-1}\right] \xrightarrow{P} \gamma, \quad \text{and} \quad \frac{1}{n}\sum_{t=1}^{n} E\left[Z_t^2|\mathcal{F}_{t-1}\right] \xrightarrow{P} c, \tag{5.3}$$

and

$$\frac{1}{n}\sum_{t=1}^{n} E\left[|Z_t|^2\mathbb{1}_{|Z_t|^2>\varepsilon\sqrt{n}}\right] \longrightarrow \gamma, \quad \text{for all } \varepsilon > 0. \tag{5.4}$$

We then have that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_t \xrightarrow{d} \mathcal{CN}_{\mathbb{C}}(0,\gamma,c). \tag{5.5}$$

Proof. It is enough to apply for each λ and μ in $\mathbb{R}$, the classical result, e.g. in Billingsley (1971), to the real MDS: $D_t = \lambda X_t + \mu Y_t$. We would deduce from the Cramér Wold Theorem a bivariate CLT:

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} \xrightarrow{d} \mathcal{N}_{\mathbb{R}^2}(0, \Sigma). \tag{5.6}$$

To show that D_t satisfies the lindeberg condition i.e. (5.4) we first remark that $|D_t| \leq \sqrt{2}\max(\mu, \lambda)|Z_k|$. We set $M = \sqrt{2}\max(\mu, \lambda)$, it follows that

$$\sum_{t=1}^{n} E\left[D_t^2 \mathbb{1}_{|Z_t|^2 > \varepsilon\sqrt{n}}\right] \leq \sum_{t=1}^{n} E\left[D_t^2 \mathbb{1}_{|Z_t|^2 > \varepsilon\sqrt{n}}\right] + E\left[D_t^2 \mathbb{1}_{|Z_t|^2 > \varepsilon\sqrt{n}}\right]$$
$$\leq$$

The entries of Σ can be deduced from the limiting variance of D_n for all λ and μ,

$$\frac{1}{n} \sum_{t=1}^{n} E\left[D_t^2 | \mathcal{F}_{t-1}\right] \xrightarrow{P} \sigma^2(\lambda, \mu) = (\lambda\ \mu)\Sigma \begin{pmatrix} \lambda \\ \mu \end{pmatrix}.$$

In other words, we have that

$$\frac{\lambda^2}{n} \sum_{t=1}^{n} E\left[X_t^2 | \mathcal{F}_{t-1}\right] + \frac{\lambda\mu}{n} \sum_{t=1}^{n} E\left[X_t Y_t | \mathcal{F}_{t-1}\right] + \frac{\mu^2}{n} \sum_{t=1}^{n} E\left[Y_t^2 | \mathcal{F}_{t-1}\right] \xrightarrow{P} \sigma^2(\lambda, \mu),$$

and those three limits can be deduced from (5.3) by using (5.1). □

The next proposition provides a complex version of Cramér Wold Theorem, under the hypothesis of tightness, the which is needed since H has infinite dimension.

Proposition 9. *Let $(Z_t)_{t\in\mathbb{Z}}$ be a tight sequence of random elements in the complex Hilbert space H. The following statements are equivalents:*

(i) $Z_n \xrightarrow{d} \mathcal{CN}_H(0, \Gamma, C)$,

(ii) $(\operatorname{Re} Z_n, \operatorname{Im} Z_n) \xrightarrow{d} \mathcal{N}_{H_0 \times H_0}(0, \Sigma)$,

(iii) for all $w = u + \mathrm{i}v \in H$, we have that

$$\operatorname{Re}\langle Z_n, w\rangle \xrightarrow{d} \mathcal{N}_{\mathbb{R}}\left(0, \langle \Sigma\tilde{w}, \tilde{w}\rangle_{H_0\times H_0}\right), \text{ where } \tilde{w} = (u, v)' \in H_0 \times H_0.$$

(iv) for all $w \in H$, we have that

$$\langle Z_n, w\rangle \xrightarrow{d} \mathcal{CN}_{\mathbb{C}}\left(0, \langle \Gamma w, w\rangle, \langle Cw, w\rangle\right).$$

If furthermore, $\sup_{n\geq 1}\|Z_n\| < \infty$, the statements (iii) and (iv) can be relaxed, i.e. it is enough to consider all w in some dense subset $\widetilde{H}$ of H.

Proof. *(i)* and *(ii)* are equivalent by definition and both imply *(iii)* and *(iv)*. the tightness provides the existence of a weak convergent subsequence, i.e. $Z_{n_k} \xrightarrow{d} Z$. Let $w = u + \mathrm{i}v \in H$. We have that $\operatorname{Re}\langle Z_n, w\rangle = \langle X_n, w\rangle + \langle Y_n, w\rangle$, and thus *(iii)* implies that the weak limit Z is Gaussian and determined by its *real* covariance operator Σ, hence it is unique. Assertion *(iv)* is just a reformulation of *(iii)*. Note that the limiting *complex* *variance* and *relation* are indeed that of $\langle Z, w\rangle$, since $E|\langle Z, w\rangle|^2 = \langle EZ\langle w, Z\rangle, w\rangle$ and $E|\langle Z, w\rangle|^2 = \langle EZ\langle w, Z\rangle, w\rangle$. The existence (and uniqueness) of theses operators follow from the polarization identities (5.2). We now assume that $\sup_{n\geq 1}\|Z_n\| < \infty$ and consider some dense subset $\widetilde{H}$ of H such that *(iii)* is satisfied. let $\varepsilon > 0$, $w \in H$ and h be a bounded Lipschitz function on H. There exists an $w' \in \widetilde{H}$ such that $\|w - \tilde{w}\| \leq \varepsilon$. It follows that

$$\begin{aligned}
&|Eh(\mathrm{Re}\langle Z_n, w\rangle) - Eh(\mathrm{Re}\langle Z, w\rangle)| \;\leq\; |Eh(\mathrm{Re}\langle Z_n, w\rangle) - h(\mathrm{Re}\langle Z_n, w'\rangle)| \\
&+ |Eh(\mathrm{Re}\langle Z_n, w'\rangle) - Eh(\mathrm{Re}\langle Z, w'\rangle)| + |Eh(\mathrm{Re}\langle Z, w'\rangle) - h(\mathrm{Re}\langle Z, w\rangle)| \\
&\leq\; \|h\|_\infty \|w - w'\| \left(E\|Z_n\| + E\|Z\|\right) + |Eh(\mathrm{Re}\langle Z_n, w\rangle) - h(\mathrm{Re}\langle Z_n, w'\rangle)| \\
&\leq\; \|h\|_\infty\, \varepsilon \left(\sup_{n\geq 1}\|Z_n\| + \|Z\|\right) + \varepsilon,
\end{aligned}$$

for some n large enough. The case of *(iv)* is similar. □

5.2 Functional moving average

In this section, we present a moments method to estimate the functional moving average parameter Θ defined in (1.6). We denote by Σ the variance operator of its innovations $(\varepsilon_t)_{t\in\mathbb{Z}}$. If we further assume that $\|\Theta\| < 1$, then $(X_t)_{t\in\mathbb{Z}}$ is invertible with respect to the innovations $(\varepsilon_t)_{t\in\mathbb{Z}}$. Indeed, we have that

$$X_t = \varepsilon_t + \Theta(X_{t-1}) - \Theta^2(X_{t-2}) + \Theta^3(X_{t-3}) \pm \dots$$

Then for some integer $p \geq 2$, there exists an H-valued stationary time series $(Y_t)_{t\in\mathbb{Z}}$ such that

$$Y_t = \varepsilon_t + \Theta(Y_{t-1}) - \Theta^2(Y_{t-2}) \pm \cdots - \Theta^p(Y_{t-p}). \tag{5.7}$$

This process is actually a functional AR(p) with parameter $\boldsymbol{\psi} = (\Theta, -\Theta^2, \dots, -\Theta^p)$, our idea is to estimate Θ through this model. To this end, We further define $C_k = E(X_k \otimes X_0)$, $D_k = E(Y_k \otimes Y_0)$,

$$\boldsymbol{C} = \begin{pmatrix} C_0 & C_1 & C_2 & \cdots & C_{p-1} \\ C_{-1} & C_0 & C_1 & \cdots & C_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{1-p} & C_{2-p} & C_{3-p} & \cdots & C_0 \end{pmatrix}$$

and $\boldsymbol{D}$ analogously with C_k replaced by D_k. The corresponding estimates from X sample and the (unobserved) Y sample are denoted by $\widehat{\boldsymbol{C}}$ and $\widehat{\boldsymbol{D}}$. All these are linear operators on H^p. If $\boldsymbol{v} = (v_1, \dots, v_p)' \in H^p$ is a vector of functions and $\boldsymbol{w} = (w_1, \dots, w_p) \in H^p$ we equip this space with inner product $\langle \boldsymbol{v}, \boldsymbol{w}\rangle = \sum_{k=1}^d \langle v_k, w_k\rangle$. This gives rise to a new Hilbert space with norm $\|\boldsymbol{v}\|^2 = \sum_{k=1}^p \|v_k\|^2$. All norms and inner products will be denoted in the same way, independent of the space. It will be clear from the context which one we mean. Note that

$$\boldsymbol{C}(\mathbf{v}) = \left(\sum_{k=0}^{p-1} C_k(v_{k+1}), \sum_{k=0}^{p-1} C_{k-1}(v_{k+1}), \dots, \sum_{k=0}^{p-1} C_{k-p+1}(v_{k+1})\right)'.$$

Let then $\boldsymbol{d} = (D_1, D_2, \ldots, D_p)$ and $\boldsymbol{c} = (C_1, C_2, \ldots, C_p)$. We denote the corresponding estimates $\widehat{\boldsymbol{d}}$ and $\widehat{\boldsymbol{c}}$. The so–called Yule–Walker equations for the model (5.7) model are

$$\begin{aligned}
D_1(v_1) &= \Theta(D_0(v_1)) \quad - \Theta^2(D_{-1}(v_1)) \;\pm \cdots - \Theta^p(D_{1-p}(v_1))\\
D_2(v_2) &= \Theta(D_1(v_2)) \quad - \Theta^2(D_0(v_2)) \quad \pm \cdots - \Theta^p(D_{2-p}(v_2))\\
\cdots &= \cdots\\
D_p(v_p) &= \Theta(D_{p-1}(v_p)) - \Theta^2(D_{p-2}(v_p)) \pm \cdots - \Theta^p(D_0(v_p))
\end{aligned}$$

and thus

$$\boldsymbol{d}(\boldsymbol{v}) := \sum_{j=1}^{p} D_j(v_j) = \boldsymbol{\psi}(\boldsymbol{D}(\boldsymbol{v})). \tag{5.8}$$

Recall that $\boldsymbol{\psi} = (\Theta, -\Theta^2, \ldots, \Theta^p) : H^p \to H$ is the true parameter of the AR(p) model (5.7). Let $\mathbf{v}_k$ be the k-th eigenfunction of $\boldsymbol{D}$ and δ_k be the corresponding eigenvalue. Then (5.8) implies that $\frac{\mathbf{d}(\mathbf{v}_k)}{\delta_k} = \boldsymbol{\psi}(\mathbf{v}_k)$, we thus define the estimator

$$\widehat{\boldsymbol{\psi}}(\boldsymbol{x}) := \sum_{k=1}^{K} \frac{\widehat{\boldsymbol{c}}(\widehat{\mathbf{v}}_k)}{\hat{\delta}_k} \langle \boldsymbol{x}, \widehat{\mathbf{v}}_k \rangle. \tag{5.9}$$

From which we deduce two possible estimators for the moving average parameter:

(5.10)

$$\widehat{\Theta}_1 = \widehat{\boldsymbol{\psi}}_1, \quad \text{and} \quad \widehat{\Theta}_2 = -\left(\sum_{k=0}^{p} \widehat{\boldsymbol{\psi}}_k^{*} \widehat{\Sigma^{-1}} \widehat{\boldsymbol{\psi}}_k\right)^{-1} \left(\sum_{k=0}^{p-1} \widehat{\boldsymbol{\psi}}_k^{*} \widehat{\Sigma^{-1}} \widehat{\boldsymbol{\psi}}_{k+1}\right), \quad \widehat{\boldsymbol{\psi}}_0 := \mathrm{id}.$$

The second is inspired by the Durbin estimator of the scalar moving average, see Durbin (1959), note that the inverses should be define on some finite dimensional subspace of H as in (1.12). The motivation comes from heuristic arguments that shows that this estimator has asymptotically the same variance as the MLE. We have not been able to prove the consistency of neither of them. Indeed, let $x \in H$ and recall that $\mathbf{v}_k = (v_{k,1}, \ldots, v_{k,p}) \in H^p$, for all $k \geq 1$, then we have

$$\begin{aligned}
\|\widehat{\boldsymbol{\psi}}_1(x) - \Theta(x)\| \leq{}& \left\| \sum_{k=1}^{K} \left(\frac{1}{\hat{\delta}_k} - \frac{1}{\delta_k}\right) \widehat{\boldsymbol{c}}(\widehat{\mathbf{v}}_k) \langle x, \widehat{v}_{k,1} \rangle \rangle \right\| + \left\| \sum_{k=1}^{K} \frac{1}{\delta_k} \langle x, \widehat{v}_{k,1} \rangle \left(\widehat{\boldsymbol{c}}(\widehat{\mathbf{v}}_k) - \boldsymbol{d}(\widehat{\mathbf{v}}_k)\right) \right\| \\
&+ \left\| \sum_{k=1}^{K} \frac{1}{\delta_k} \langle x, \widehat{v}_{k,1} \rangle \boldsymbol{d}(\widehat{\mathbf{v}}_k - \mathbf{v}_k) \right\| \\
&+ \left\| \sum_{k=1}^{K} \frac{1}{\delta_k} \langle x, \widehat{v}_k - v_{k,1} \rangle \boldsymbol{d}(\mathbf{v}_k) \right\| + \left\| \sum_{k=1}^{K} \frac{\boldsymbol{d}(\mathbf{v}_k)}{\delta_k} \langle x, v_{k,1} \rangle - \Theta(x) \right\|.
\end{aligned}$$

Let then $K = K_n$, be a sequence that converges to $+\infty$. Under similar assumption than in Bosq (2000)[Chap. 8] and using Lemma 16 we can show that the first four terms converges to zero in probability. For the last term, we have that

$$\sum_{k=1}^{K} \frac{\boldsymbol{d}(\mathbf{v}_k)}{\delta_k} \langle x, v_{k,1} \rangle = \sum_{k=1}^{K} \sum_{j=1}^{p} (-1)^{j-1} \Theta^j(v_{k,j}) \langle x, v_{k,1} \rangle = \sum_{j=1}^{p} (-1)^{j-1} \Theta^j \left(\sum_{k=1}^{K} v_{k,j} \langle x, v_{k,1} \rangle\right).$$

But we are not able to say something else. Although $(\mathbf{v}_k)_{k\geq 1}$ is an orthonormal basis of H^p, for each j, $(v_{k,j})_{k\geq 1}$ are neither orthonormal nor basis of H.

Lemma 16. *If* $\|\Theta\| < 1$ *and* $p = p_n \nearrow +\infty$, *we have that* $\|X_t - Y_t\| = O_P(n^{-1})$.

Proof of Lemma 16. For simplicity, we assume that p is even. Note that $(Y_t)_{t\in\mathbb{Z}}$ has the following functional MA(∞) representation

$$Y_t = \varepsilon_t + \Theta(Z_{t-1}) + \sum_{k\geq 1}(-1)^k \left(\Theta^{k(p+1)}(\varepsilon_{t-k(p+1)}) + \Theta^{k(p+1)+1}(\varepsilon_{t-k(p+1)-1})\right).$$

and since $X_t = \varepsilon_t + \Theta(Z_{t-1})$, we get that

$$\|Y_t - X_t\| \leq \sum_{k\geq 1} \|\Theta\|^{k(p+1)} \|\varepsilon_{t-k(p+1)}\| + \sum_{k\geq 1} \|\Theta\|^{k(p+1)+1} \|Z_{t-k(p+1)-1}\|,$$

where

$$\sum_{k\geq 1} \|\Theta\|^{k(p+1)} \|\varepsilon_{t-k(p+1)}\| = \|\Theta\|^{p+1} \sum_{k\geq 1} \left(\|\Theta\|^{(p+1)}\right)^{k-1} \|\varepsilon_{t-k(p+1)}\| = O_P(\|\Theta\|^{p+1}).$$

□

5.3 Appendix to Chapter 4

5.3.1 Independence of random operators

In Proposition 5, we used the fact that the expectation of the product of independent operators factorizes. Since this simple fact and other basic properties of the expectation of random operators are usually not discussed in details, we provided it in the next lemma.

Lemma 17. *Let* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *be two independent random elements of* $\mathcal{L}(H)$ *such that* $E\|\boldsymbol{X}\| < \infty$ *and* $E\|\boldsymbol{Y}\| < \infty$. *We have that:*

(i) $E[\boldsymbol{X}](\phi) = E[\boldsymbol{X}(\phi)]$, *for all* $\phi \in H$;

(ii) $E[\boldsymbol{X}^*] = E[\boldsymbol{X}]^*$;

(iii) $E[\boldsymbol{X}\boldsymbol{Y}] = E[\boldsymbol{X}]E[\boldsymbol{Y}]$.

Proof. Recall that $\mathcal{L}(H)$ is a Banach space for the operator norm. Therefore, for any random operator $\boldsymbol{X}$ such that $E\|\boldsymbol{X}\| < \infty$ there exists a unique operator denoted by $E[\boldsymbol{X}]$ that satisfies $E[\xi(\boldsymbol{X})] = \xi(E[\boldsymbol{X}])$ for all $\xi \in \mathcal{L}(H)^*$, for more details see e.g. Bosq (2000). Note that $\mathbf{x} \mapsto \langle \mathbf{x}(\phi), \psi\rangle$ belongs to $\mathcal{L}(H)^*$. The definition of $E[\boldsymbol{X}]$ thus imply that $\langle E[\boldsymbol{X}](\phi), \psi\rangle = E\langle \boldsymbol{X}(\phi), \psi\rangle$ and by definition of $E[\boldsymbol{X}(\phi)]$ we have that $\langle E[\boldsymbol{X}(\phi)], \psi\rangle = E\langle \boldsymbol{X}(\phi), \psi\rangle$. Since it holds for all $\psi \in H$ we can deduce *(i)*. Similarly, the definition of $E[\boldsymbol{X}]$ and $E[\boldsymbol{X}^*]$ provides that $\langle E[\boldsymbol{X}](\phi), \psi\rangle = E\langle \boldsymbol{X}(\phi), \psi\rangle = E\langle \phi, \boldsymbol{X}^*(\psi)\rangle = \langle \phi, E[\boldsymbol{X}^*](\psi)\rangle$, this shows *(ii)*. Using *(i)* and *(ii)* we get that

$$\langle E[\boldsymbol{X}]E[\boldsymbol{Y}](\phi), \psi\rangle = \langle E[\boldsymbol{Y}](\phi), E[\boldsymbol{X}]^*(\psi)\rangle = \langle E[\boldsymbol{Y}(\phi)], E[\boldsymbol{X}^*(\psi)]\rangle = E\langle \boldsymbol{Y}(\phi), \boldsymbol{X}^*(\psi)\rangle,$$

where in the equality we used the fact that $\boldsymbol{X}(\phi)$ and $\boldsymbol{Y}(\psi)$ are independent random elements in H. To conclude we use the definition of $E[\boldsymbol{X}\boldsymbol{Y}]$ to get that $\langle E[\boldsymbol{X}\boldsymbol{Y}](\phi), \psi\rangle = E\langle \boldsymbol{X}\boldsymbol{Y}(\phi), \psi\rangle$. It then holds that $\langle E[\boldsymbol{X}]E[\boldsymbol{Y}](\phi), \psi\rangle = \langle E[\boldsymbol{X}\boldsymbol{Y}](\phi), \psi\rangle$ for all ϕ and ψ in H, which proves *(iii)*. □

5.3.2 Sufficient condition

The next lemma shows the relation between assumptions (4.15) and Assumption **A8**.

Lemma 18. *If $M < \infty$, $p = q = 1$, and $\mathfrak{A}_0$ is invertible then* (4.15) *implies* **A8**.

Proof. For any element $u \in \operatorname{span}\{\varphi_m,\, 1 \le m \le M\}$ we define it's coefficient vector $\underline{u}$ such that $u = \sum_{m=1}^{M} \underline{u}_m \varphi_m$ and the vector of scalar products $\overline{u} = (\langle u, \varphi_1\rangle, \dots \langle u, \varphi_M\rangle)^\top$. Similarly, for any operator in $\boldsymbol{\gamma} \in \operatorname{span}\{\varphi_m \otimes \varphi_{m'},\, 1 \le m, m' \le M\}$ we define the matrices $\underline{\boldsymbol{\gamma}}$ and $\overline{\boldsymbol{\gamma}}$ as follows

$$\boldsymbol{\gamma} = \sum_{m,m'=1}^{M} \underline{\boldsymbol{\gamma}}_{m,m'} \varphi_m \otimes \varphi_{m'} \quad \text{and} \quad \overline{\boldsymbol{\gamma}} = \begin{pmatrix} \langle \boldsymbol{\gamma}(\varphi_1), \varphi_1\rangle & \dots & \langle \boldsymbol{\gamma}(\varphi_M), \varphi_1\rangle \\ \vdots & \ddots & \\ \langle \boldsymbol{\gamma}(\varphi_1), \varphi_M\rangle & & \langle \boldsymbol{\gamma}(\varphi_M), \varphi_M\rangle \end{pmatrix}.$$

Actually we can prove that

$$\begin{gathered} \delta_\theta \in \operatorname{span}\{\varphi_m,\, 1 \le m \le M\} \quad \Longrightarrow \quad \mathbf{A8}(a) \\ \boldsymbol{\alpha}_\theta \in \operatorname{span}\{\varphi_m \otimes \varphi_{m'},\, 1 \le m, m' \le M\} \quad \Longrightarrow \quad \mathbf{A8}(b) \\ \boldsymbol{\alpha}_0, \boldsymbol{\beta}_\theta \in \operatorname{span}\{\varphi_m \otimes \varphi_{m'},\, 1 \le m, m' \le M\},\ \underline{\boldsymbol{\alpha}_0} \in GL(\mathbb{R}^M) \quad \Longrightarrow \quad \mathbf{A8}(c). \end{gathered}$$

To this end, first note that $\overline{u} = \Phi \underline{u}$, $\overline{\boldsymbol{\gamma}} = \Phi \underline{\boldsymbol{\gamma}} \Phi$, and $\underline{\boldsymbol{\gamma}_1 \circ \boldsymbol{\gamma}_2} = \underline{\boldsymbol{\gamma}_1} \Phi \underline{\boldsymbol{\gamma}_2}$, where Φ is the Gram-matrix of the instrument functions the which can be assumed to be linearly independent, and thus Φ is invertible. Now if we take $u = \delta_0 - \delta_\theta$ with $\delta_0, \delta_\theta \in \operatorname{span}\{\varphi_m,\, 1 \le m \le M\}$, it is clear that $\overline{u} = 0$ implies that $\underline{u} = 0$ and the first implication is proved. Similarly, if we take $\boldsymbol{\gamma} = \boldsymbol{\alpha}_0^* - \boldsymbol{\alpha}_\theta^*$ with $\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_\theta \in \operatorname{span}\{\varphi_m \otimes \varphi_{m'},\, 1 \le m, m' \le M\}$, it is clear that $\overline{\boldsymbol{\gamma}} = 0$ implies that $\underline{\boldsymbol{\gamma}} = 0$ and the second implication is proved. Finally to prove the third implication, we take $\boldsymbol{\gamma} = \boldsymbol{\alpha}_0^* \circ (\boldsymbol{\beta}_0^* - \boldsymbol{\beta}_\theta^*)$, then $\overline{\boldsymbol{\gamma}} = \Phi \underline{\boldsymbol{\alpha}_0^*} \Phi \underline{(\boldsymbol{\beta}_0^* - \boldsymbol{\beta}_\theta^*)} \Phi$, to conclude we need that $\underline{\boldsymbol{\alpha}_0}$ is invertible. This follows from the assumption since $\underline{\boldsymbol{\alpha}_0} = \Phi^{-1} \overline{\boldsymbol{\alpha}_0}$ and $\overline{\boldsymbol{\alpha}_0} = \mathfrak{A}_0$. ☐

5.3.3 Discussion on positivity constraints

To prove Theorems 13 and 14 we only need the following positivity constraints

$$\delta_\theta \in H_*^+,\ \boldsymbol{\alpha}_i, \boldsymbol{\beta}_j \in \mathcal{L}^+(H),\ \forall \theta \in \Theta, \tag{5.11}$$

and that, componentwise

$$\mathfrak{d} = \Phi d > 0,\ \mathfrak{A}_i = \Phi A_i \ge 0,\ \text{and}\ \ \mathfrak{B}_j = \Phi B_j \ge 0, \tag{5.12}$$

for $i = 1, \dots, q$ and $j = 1, \dots, p$. Condition (5.11) is needed to define our functional GARCH model (4.1)–(4.2) and condition (5.12) is needed when we use the auxiliary CCC-GARCH process defined in the proofs of Theorems 13 and 14. However, equation (5.11) is not explicit and cannot be used in practice. Note that since the functions $\varphi_1, \dots, \varphi_M$ are non-negative, the assumption that $d \ge 0$, $A_i \ge 0$ and $B_j \ge 0$ in (4.15) is stronger than (5.11). In other words, the parameter space used in Section 4.3.1 is too small. On the other hand equation (5.12) alone does not imply (5.11), i.e. there is some non-negative $\mathfrak{d}$, $\mathfrak{A}_i$ and $\mathfrak{B}_j$ for which the corresponding δ is not in H_*^+, $\boldsymbol{\alpha}_i$ or $\boldsymbol{\beta}_j$ is not in $\mathcal{L}^+(H)$, and thus, equation (5.12) does not define a proper parameter space, it's too big. However, since $\widehat{\theta}_n$ converges to the true value of the parameter, which satisfies both positivity constraints (5.11) and (5.12), we might still enlarge the parameter space by only imposing (5.12).

Conclusion

In the first two chapters, we have obtained asymptotic results for the spectral analysis of functional time series. In Chapter 2, we have provided a CLT for the discrete Fourier transform of a functional time series under very general conditions. Whereas in Chapter 3, we have shown that the maximum over all fundamental frequencies of the norm of the discrete Fourier transform is in the attraction domain of the Gumbel distribution. This last result can be used to construct various tests, for detecting periodic patterns in a functional or multivariate time series. In Section 3.3.1, we have considered such a test for a non i.i.d. multivariate process. It would be interesting to extend this test to the functional setting. In a second step, we might also extend it to a much more general notion of dependence, such as the ones we considered in Chapter 2.

In Chapter 4 we have generalized the GARCH(p,q) model to the functional setting. We have provided a sharp sufficient condition for the existence of a stationary functional GARCH process with given coefficients. So far we do not know if this condition is necessary, this could be interesting to investigate this question, in order e.g. to make inference on the model. We have proposed an estimation method inspired by the quasi–maximum likelihood method. As we mentioned in Section 5.3.3, one drawback of this method is related to the positivity constraints of our model, which are not straightforward when using projecting functional data. To solve this problem, an alternative model without positivity constraints on the coefficients, such as the log–GARCH, could be investigated. It might also be interesting to investigate more precisely the role played by choice of the functions $\varphi_1, \ldots, \varphi_M$ used in our procedure. Various extensions of the model could be considered, e.g. multivariate model for modelling several functional GARCH processes jointly, or Markov–switching GARCH models.

References

AUE, A., DUBART NORINHO, D. AND HÖRMANN, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110, 378–392.

AUE, A., HORVÁTH, L. AND PELLATT., D. F. (2016). Functional generalized autoregressive conditional heteroskedasticity. *Journal of Time Series Analysis*, 38, 3–21.

AUE, A. AND KLEPSCH, J. (2017). Estimating invertible functional time series. *Functional Statistics and Related Fields*, 51–58.

AUE, A. AND KLEPSCH, J. (2017). Estimating functional time series by moving average model fitting. arXiv preprint `arXiv:1701.00770`.

BAGCHI, P., CHARACIEJUS, V. AND DETTE, H. (2018). A simple test for white noise in functional time series. Journal of Time Series Analysis, 39(1), pp.54-74.

BERKES, I., HORVÁTH, L. AND RICE, G. (2013). *Weak invariance principles for sums of dependent random functions.* Stochastic Processes and their Applications, 123, 385–403.

BILLINGSLEY, P.(1971). *Convergence of Probability Measures.* First edition, Wiley series in probability and mathematical statistics, New York.

BOLLERSLEV, T.(1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307–327.

BOSQ, D. (2000). *Linear Processes in Function Spaces.* Springer, New York.

BOUGEROL, P. AND PICARD, N. (1992). Strict stationarity of generalized autoregressive processes. *The Annals of Probability*, 20, 1714–1730.

BRADLEY, R.C. (2005). *Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions.* Probability Surveys Vol. 2, 107–144.

BREZIS, H. (2010). *Functional Analysis, Sobolev Spaces and Partial Differential Equations.* Springer 1st Edition.

BRILLINGER, D.R. (1969): Asymptotic properties of spectral estimates of second order. Biometrika, 56, 2, p. 375 Printed in Great Britain

BROCKWELL, P.J. AND DAVIS, R.A. (1991). Time Series: Theory and Methods. Springer, New York.

CEROVECKI, C. FRANCQ, C., HÖRMANN, S. AND ZAKOIAN, J.–M. (2018): Functional GARCH models: the quasi-likelihood approach and its applications. MPRA preprint *mpra.ub.uni-muenchen.de/83990/.*

CEROVECKI, C. AND HÖRMANN, S. (2017). On the CLT for discrete Fourier transforms of functional time series. *Journal of Multivariate Analysis*, 154, 282–295.

CHATTERJI, S.D. (1964). *A note on the convergence of Banach-space valued martingales.* Mathematische Annalen, Volume 153, Issue 2, pp 142–149.

CHERNOZHUKOV, V., CHETVERIKOV, D. AND KATO, K. (2017): Central limit theorems and bootstrap in high dimensions. The Annals of Probability, 45(4), pp.2309–2352.

CUEVAS, A. (2014). *A partial overview of the theory of statistics with functional data.* Journal of Statistical Planning and Inference 147, 1-23.

DAMON, J., GUILLAS, S. (2005). *Estimation and simulation of autoregressive Hilbertian processes with exogenous variables.* Statistical Inference for Stochastic Processes, 8, 185–204.

DAVIS, R.A. AND MIKOSCH, T. (1999) The maximum of the periodogram of a non-Gaussian sequence. The Annals of Probability, 27(1), pp.522-536.

DIDERICKSEN, D., KOKOSZKA, P., AND ZHANG, X. (2012). *Empirical properties of forecasts with the functional autoregressive model.* Computational Statistics, 27, 285–298.

DURBIN, J. (1959). *Efficient estimation of parameters in moving-average models.* Biometrika, 306-316.

EICHLER, M. AND VAN DELFT, A. (2017). Locally stationary functional time series. *arXiv preprint* 1602.05125v2.

ENGLE., R. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50, 987–1008.

ESCANCIANO, J., C. (2009). Quasi-Maximum Likelihood Estimation of Semi-Strong GARCH Models. *Econometric Theory*, 25, 561–570

FERRATY, F., GOIA, A., AND VIEU, P. (2002). *Functional nonparametric model for time series: a fractal approach for dimension reduction.* Test 11, 317–344.

FERRATY, F. AND VIEU, P. (2006). *Nonparametric functional data analysis: theory and practice.* Springer Science & Business Media.

FRANCQ, C. AND ZAKOÏAN, J. M. (2011). *GARCH models: structure, statistical inference and financial applications.* John Wiley & Sons.

GOIA, A. AND VIEU, P. (2016). *An introduction to recent advances in high/infinite dimensional statistics.* Journal of Multivariate Analysis 146 (2016): 1–6.

GUILLAS, S., BAKARE, A., MORLEY, J., SIMONS, R. (2011). *Functional autoregressive forecasting of long-term seabed evolution.* Journal of Coastal Conservation, 15, 337–351.

HÖRMANN, S., HORVÁTH, L., AND REEDER, R. (2013). *A functional version of the ARCH model.* Econometric Theory, 29(02), 267-288.

HÖRMANN, S., KIDZIŃSKI, L. AND HALLIN, M. (2015). *Dynamic Functional Principal Component.* Journal of the Royal Statistical Society: Series B, 77, 319–348

Hörmann, S., Kidziński, L. and Kokoszka, P. (2014). *Estimation in functional lagged regression.* Journal of Time Series Analysis, 36, 541–561.

Hörmann, S., Kokoszka, P. (2010). *Weakly dependent functional data.* The Annals of Statistics, 38, 1845–1884.

Hörmann, S., Kokoszka, P. and Nisol, G. (2017). *Testing for periodicity in functional time series.* The annals of statistics.

Horváth, L., Kokoszka, P. (2014). Inference for functional data with applications (Vol. 200). Springer Science & Business Media.

Horváth, L. Kokoszka, P. and Reeder, R. (2013). *Estimation of the mean of functional time series and a two-sample problem.* JRSSB, 75.1, 103–122.

Horváth, L., Kokoszka, P. and Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics*, 179, 66–82.

Horváth, L., Rice, G., and Whipple, S. (2016). *Adaptive bandwidth selection in the estimation of the long run covariance of functional time series.* Computational Statistics and Data Analysis, 100, 676–693.

Hsing, T., and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators.* John Wiley & Sons.

Hunt, R.A. and Young, W.S. (1974). A weighted norm inequality for Fourier series. Bulletin of the American Mathematical Society, 80(2), pp.274-277.

Hyndman, R.J. and Shang, H.L. (2009). *Forecasting functional time series.* Journal of the Korean Statistical Society, 38, No. 3, 199–211.

Jirak, M. (2013). *On weak invariance principles for sums of dependent random functionals.* Statistics and Probability Letters, 83, 2291–2296.

Kang, S. and Serfozo, R.F. (1999) Extreme values of phase-type and mixed random variables with parallel-processing examples. Journal of applied probability, 36(1), pp.194-210.

Kargin, V. and Onatski, A. (2008). *Curve forecasting by functional autoregression.* Journal of Multivariate Analysis, 99, 2508–2526.

Karhunen, K. (1947). Über lineare Methoden in der Wahrscheinlichkeitsrechnung (Vol. 37). Universitat Helsinki.

Kingman, J. F. C. (1973). Subadditive ergodic theory. *The Annals of Probability*, 1, 883–899.

Klepsch, J. and Klüppelberg, C. (2017). An innovations algorithm for the prediction of functional linear processes. *Journal of Multivariate Analysis*, 155, 252–271.

Kokoszka, P. and Mikosch, T. (2000). *The periodogram at the Fourier frequencies.* Stochastic Processes and their Applications, 86, 49–79.

Kokoszka, P., Rice, G. and Shang, H.L. (2017) Inference for the autocovariance of a functional time series under conditional heteroscedasticity. Journal of Multivariate Analysis, 162, pp.32-50.

Kowal, D.R., Matteson, D.S., and Ruppert, D. (2016). *Functional autoregression for sparsely sampled data.* arXiv:1603.02982, working paper.

Ledoux, M. and Talagrand, M. (1991) Probability in Banach Spaces. Springer-Verlag, Berlin, Heidelberg.

Liebl, D. (2013). *Modeling and forecasting electricity prices: A functional data perspective.* The Annals of Applied Statistics, 7, 1562–1592.

Lin, Z. and Liu, W. (2009). On maxima of periodograms of stationary processes. The Annals of Statistics, pp.2676-2695.

Mas, A. (2007). *Weak convergence in the functional autoregressive model.* Journal of Multivariate Analysis, 98, 1231–1261.

Merlevède, F., Peligrad, M. and Utev, S. (1997). *Sharp Conditions for the CLT of Linear Processes in a Hilbert Space.* Journal of Theoretical Probability, 10, No. 3, 681–693.

Mikosch, T. and Straumann, D. (2002) Whittle estimation in a heavy-tailed GARCH (1, 1) model. Stochastic processes and their applications, 100(1-2), pp.187-222.

Mikusiński, J. (1978). *The Bochner Integral.* Birkhauser, Basel.

Nazarov, F. (2003) On the Maximal Perimeter of a Convex Set in $\mathbb{R}^n$ with Respect to a Gaussian Measure. In Geometric aspects of functional analysis (pp. 169-187). Springer, Berlin, Heidelberg.

Panaretos, V.M. and Tavakoli, S. (2013). *Fourier analysis of stationonary time series in function spaces.* The Annals of Statistics, 41, No. 2, 568–603.

Panaretos, V.M. and Tavakoli, S. (2013). *Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series.* Stochastic Processes and their Applications, 123, 2779–2807.

Paparoditis, E. (2016) *Sieve Bootstrap for Functional Time Series.* arXiv:1609.06029, working paper.

Paparoditis, E. (2017). Sieve bootstrap for functional time series. *arXiv preprint* 1609.06029v2.

Parthasarathy, K. R. (1967). (1967) *Probability measures on metric spaces.* Academic press, New York.

Peligrad, M. and Wu, W.B. (2010). *Central limit theorem for Fourier transforms of stationary processes.* The Annals of Probability, 38, No. 5, 2009–2022.

Picinbono, B. (1996) Second-order complex random vectors and normal distributions. IEEE Transactions on Signal Processing, 44(10), pp.2637-2640.

Politis, D.N. (2011) Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices. Econometric Theory, 27(4), pp.703-744.

Pötscher, B.M. and Prucha, I.R. (1997). *Dynamic Nonlinear Econometric Models.* Springer.

Pruitt, W. E.(1966), *Summability of independent random variables.* Journal of Mathematical and Mechanics 15: 769-776.

Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis.* Springer, New York, second edition.

Račkauskas, A. and Suquet, Ch. (2010). *On limit theorems for Banach-space-valued linear processes.* Lithuanian Mathematical Journal, 50, No. 1, 71–87.

Rosenthal, H.P. (1970) On the subspaces of L^p (p¿2) spanned by sequences of independent random variables. Israel Journal of Mathematics, 8(3), pp.273-303.

Ruiz-Medina, M. D., Romano, E., and Fernàndez-Pascual, R. (2016). *Plug-in prediction intervals for a special class of standard ARH(1) processes.* Journal of Multivariate Analysis, 146, 138–150.

Turbillon, C. (2007). *Estimation et prévision des processus Moyenne Mobile fonctionnels.* Thèse de doctorat dirigée par Denis Bosq, Jean-Marie Marion et Besnik Pumo. Université Paris VI.

Walker, A.M. (1965) Some asymptotic results for the periodogram of a stationary time series. Journal of the Australian Mathematical Society, 5(1), pp.107-128.

Wu, W.B. (2004). *Fourier transforms of stationary processes.* Proceeding of the American society, 133, 285–293.

Zhu, T. and Politis, D.N. (2017). Kernel estimation of first-order nonparametric functional autoregression model and its bootstrap approximation. *Electronic Journal of Statistics*, forthcoming.

Inférence asymptotique pour des processus stationnaires fonctionnels

Mots–clés : *Données fonctionnelles, Modèles conditionnellement hétéroscédastiques, Quasi-maximum de vraisemblance, Séries temporelles, Statistiques des extrêmes, Test de périodicité, Théorème central limite, Transformée de Fourier.*

Nous abordons divers problèmes concernant les séries temporelles fonctionnelles. Il s'agit de processus stochastiques discrets à valeurs dans un espace fonctionnel. La principale motivation provient de l'interprétation séquentielle d'un phénomène continu. Si par exemple on observe des données météorologiques au cours du temps de manière continue, il est naturel de segmenter ce processus en une série temporelle fonctionnelle indexée par les jours. Chaque terme de la série représente la courbe journalière. Dans un premier temps, nous nous sommes intéressés à l'analyse spectrale. Plus précisément nous avons montré que sous des hypothèses très générales, la transformée de Fourier discrète d'une telle série est asymptotiquement normale et a pour variance l'opérateur de densité spectrale. Une application possible de ce résultat est de tester la présence de composantes périodiques dans une série fonctionnelle. Nous avons développé un test valable pour une fréquence arbitraire. Pour ce faire, nous avons étudié le comportement asymptotique du maximum de la norme de la transformée de Fourier. Enfin, nous avons travaillé sur la généralisation fonctionnelle du modèle GARCH. Ce modèle permet de décrire la dynamique de la volatilité, c'est-à-dire de la variance conditionnelle, dans les données financières. Nous avons proposé une méthode d'estimation des paramètres du modèle, inspirée de l'estimateur de quasi-maximum de vraisemblance. Nous avons montré que cet estimateur est convergent et asymptotiquement normal, puis nous l'avons évalué sur des simulations et appliqué à des données réelles.

Asymptotic inference in stationary functional processes

Keywords: *Central limit theorem, Conditional heteroscedastic models, Functional data analysis, Fourier transform, Periodicity test, Quasi–maximum likelihood, Statistics for extreme values, Time series.*

In this thesis we address some issues related to functional time series, which consists in a discrete stochastic process valued in a functional space. The main motivation comes from a sequential approach of a continuous phenomenon. For example, if we observe some meteorological data continuously over time, then it is natural to segment this process into a functional series indexed by days, each term representing the daily curve. The first part is devoted to spectral analysis, more precisely we study the asymptotic behavior of the discrete Fourier transform. We show that, under very general conditions, the latter is asymptotically normal, with variance equal to the spectral density operator. An application of this result is the detection of periodic patterns in a functional time series. We develop a test to detect such patterns, which is valid for an arbitrary frequency. We show that the asymptotic distribution of the norm of the discrete Fourier transform belongs to the attraction domain of the Gumbel distribution. In a second part, we work on the functional generalization of the GARCH model. This model is used to describe the dynamics of volatility, i.e. conditional variance, in financial data. We propose an estimation method inspired by the quasi-maximum likelihood estimator, although the proper likelihood function does not exist in infinite dimension. We show that this estimator is convergent, asymptotic normal and we evaluate its performances on simulated and real data.

