

Analysis of Distributional Dynamics for Repeated Cross-Sectional and Intra-Period Observations*

Bo Hu

Institute of New Structural Economics
Peking University

Joon Y. Park

Department of Economics
Indiana University
and Sungkyunkwan University

Junhui Qian

School of Economics
Shanghai Jiao Tong University

February 20, 2020

Abstract

This paper introduces a novel approach to investigate the dynamics of state distributions, which accommodate both cross-sectional distributions of repeated panels and intra-period distributions of a time series observed at high frequency. In our approach, densities of the state distributions are regarded as functional elements in a Hilbert space, and are assumed to follow a functional autoregressive model. We propose an estimator for the autoregressive operator, establish its consistency, and provide tools and asymptotics to analyze the forecast of state density and the moment dynamics of state distributions. We apply our methodology to study the time series of distributions of the GBP/USD exchange rate intra-month returns and the time series of cross-sectional distributions of the NYSE stocks monthly returns. Finally, we conduct simulations to evaluate the density forecasts based on our model.

JEL classification codes: C13, C14, C22

Key words and phrases: distributional dynamics, repeated panels, intra-period distributions, functional autoregression, time-varying density, density forecasting

*We are grateful for useful comments to Yoosoon Chang, Chang Sik Kim, Peter Phillips, Marine Carrasco and the participants of 2015 Midwest Econometrics Group Meeting, 2016 SETA, Time Series Workshop on Macro and Financial Economics, International Symposium on Financial Engineering and Risk Management, SoFiE, CMES, and the attendants of the seminars at Toulouse School of Economics, Einaudi Institute of Economics and Finance, University of Cincinnati, University of Montreal and Indiana University. Junhui Qian would like to thank National Natural Science Foundation of China for financial support (Project ID: 71673183).

1. Introduction

Many important issues in economics are related to the time evolution of state distributions of economic variables, which are defined to be either cross-sectional or intra-period distributions of the underlying economic variables. For example, the time series of cross-sectional or intra-period distributions of returns from traded assets contain important information on the dynamics of risk and return in financial markets. However, tracking the complete distributional dynamics is not an easy task because of the infinite dimensional nature of state distributions. The conventional approaches usually get around the problem by only considering some particular dimensions of the distributions such as the mean, the variance, or some specific quantiles. Information beyond these dimensions, however, would be necessarily ignored. In this paper, we develop a novel analytical framework that can be used to track and exploit the dynamics of entire state distributions.

We consider the functional autoregressive (FAR) model of time-varying distributions. Allowing the distributions to be infinite-dimensional, we adopt a nonparametric approach to modeling distributional dynamics. The parametric approach has been well studied, including the autoregressive conditional moment models by Engle (1982), Bollerslev (1986), Engle et al. (1987), Harvey and Siddique (1999) and Brooks et al. (2005), the autoregressive conditional density model by Hansen (1994), and the autoregressive quantile models by Koenker and Xiao (2006) and Xiao and Koenker (2009), among many others. Our approach, on the contrary, does not require any parametric assumptions on the underlying state distributions. Nevertheless, it allows us to investigate full temporal dynamics of state distributions, including dependence structures among moments and tail probabilities, within a single framework.

For our analysis, we represent state distributions by their density functions and regard them as random elements of a Hilbert space. Of course, it is also possible to use other functions such as distribution functions and quantile functions to represent state distributions.¹ However, we choose density functions, since they provide clear and direct interpretations on the dynamics of moments, tail probabilities and other important features of state distributions. Indeed, in our framework, the cross-sectional moments and tail probabilities are defined simply as the inner products of state density functions with some deterministic functions in a Hilbert space. No other functions representing state distributions have this nice property. We fit our FAR model using the time series of state density functions, which

¹van den Boogaart et al. (2014) propose to use the so-called Bayes Hilbert space, which uses log-densities instead of densities themselves, to study distributional dynamics. Though their approach has some advantages, it does not allow for any of the methodologies developed in our paper to analyze distributional dynamics.

are obtained from repeated cross-sectional and intra-period high frequency observations.² Once fitted, we estimate the leading progressive/regressive features and conduct various impulse response analyses to reveal the way in which future state distributions respond to changes in past state distributions. In addition, we develop a variance decomposition method that relates a specific feature of a current state distribution to the moments of its past. Lastly, we propose a density predictor and find its asymptotic properties.

Recently, there has been a rapid development in the theory and applications of functional time series analysis. Many authors, including Besse et al. (2000), Bosq (2000), Antoniadis et al. (2006), Ferraty and Vieu (2006), Mas (2007), Kargin and Onatski (2008), Hörmann and Kokoszka (2010), Horváth et al. (2010), Didericksen et al. (2012), Kokoszka and Reimherr (2013), Panaretos and Tavakoli (2013b), Panaretos and Tavakoli (2013a), Aue et al. (2015), Aue et al. (2017) and Klepsch and Klüppelberg (2017), among others, studied various aspects of functional data. In many of these works, however, temporal dependencies in the functional dynamics are used in a relatively implicit way to make forecasts. We in this paper emphasize the investigation of the functional dynamics themselves in the distributional dynamics setting, and develop tools to interpret them in economically meaningful manners.

We apply our methodology to two representative state distributions: the *intra-month* distributions of the UK Pound/US Dollar (GBP/USD) exchange rate 15-minute log returns and the *cross-sectional* distributions of the New York Stock Exchange (NYSE) stocks monthly returns. In both cases, we find that observations/stocks with small returns play the most important role in interacting with the past distribution and in determining the future distribution. Outliers turn out to be unimportant in the distributional dynamics. We also find that various characteristics such as the moments and the tail probabilities of the current distribution respond differently to shocks to, and subsequent changes in, the relative frequency of observations with different levels of returns. The momentum effect appears in the dynamics of the first moment: The more observations with positive (negative) returns in the past distribution, the higher (lower) the first moment of the current distribution. On the other hand, sizes of the returns play a more important role in the dynamics of the second moment; the more observations with small (large) returns in the past distribution, the lower (higher) the second moment of the current distribution.

In contrast to the symmetry found in the behaviors of left and right tail probabilities in the foreign exchange rate return application, those in the stock return application appear

²In all our applications, the number of cross-sectional or intra-period observations is relatively much bigger than the number of time series observations. Therefore, we ignore the statistical errors incurred by the estimation of state densities using repeated cross-sectional and intra-period observations.

to be quite asymmetric. In the former, a higher relative frequency of large returns in the past distribution, regardless of the signs of these returns, increases both the current left and right tail probabilities rather symmetrically. In the latter, however, the current left tail probability is mostly determined by the relative frequency of moderate-to-large negative returns in the past distribution, whereas the current right tail probability is positively affected both by the relative frequency of moderate-to-large negative returns, though to a lesser degree, and by the relative frequency of positive returns in the past distribution. The variance decomposition exercises indicate that in both cases, the variances of the current first moments are not explained much by the variances of the past integral moments, while the current second moments are more closely connected to the past second moments. These results lend support to the usual observation that in financial markets, the mean is not quite predictable while the volatility is much more persistent, and thus predictable. These moment dependence findings are also consistent with the rolling out-of-sample forecast results using our model.

The rest of the paper is organized as follows. In Section 2, we introduce the functional autoregressive model of time-varying densities and the necessary theoretical background for understanding the model. Section 3 explains how we may estimate the model and develops the relevant statistical theory. In particular, our estimators of the autoregressive operator and of the error variance operator are shown to be consistent under appropriate regularity conditions. Section 4 presents density forecast along with its asymptotic theory, and provides a variance decomposition method that is useful in studying the moment dynamics of state distributions. In Section 5, we apply our model and methodology to the state distributions representing intra-month distributions of high frequency GBP/USD exchange rate returns and cross-sectional distributions of the NYSE stocks monthly returns. Section 6 summarizes our simulation results on the finite sample performance of the density predictor. Section 7 concludes the paper. All mathematical proofs are collected in Appendix.

2. The Model and Preliminaries

In this section, we present our model and some preliminaries that are necessary for the development of our theory and methodology in the paper.

2.1. The Model

In this paper, we consider a random sequence (f_t) of probability densities of state distributions, which represent some distributional aspect of an economy, for time $t = 1, 2, \dots$. The density functions (f_t) are time varying, and are regarded as random elements taking values

in the Hilbert space $L^2(C)$ of square integrable functions on some compact subset C of \mathbb{R} or \mathbb{R} itself. We may introduce the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and more formally define $f_t : \Omega \rightarrow L^2(C)$ for $t = 1, 2, \dots$ with the required measurability condition.

We assume that (f_t) has a well defined common mean $\mathbb{E}f$, whose precise meaning will be introduced later in Section 2.2. Moreover, we let the demeaned state densities (w_t) , $w_t = f_t - \mathbb{E}f$, be generated according to

$$w_t = Aw_{t-1} + \varepsilon_t \tag{1}$$

for $t = 1, 2, \dots$, where A is a bounded linear operator, called the autoregressive operator, and (ε_t) is a functional white noise process.³ The meaning of a functional white noise process will be introduced later in Section 2.2. Note that (w_t) takes values in a subset H of $L^2(C)$, which is given by⁴

$$H = \{v \in L^2(C) \mid \langle 1, v \rangle = 0\},$$

where the inner product in the Hilbert space H is defined by $\langle z, w \rangle = \int_C z(x)w(x)dx$, and 1 denotes the constant function taking value 1 everywhere on the support. Throughout the paper, we assume that A is an operator on H , so that (ε_t) also take values in H .

The model introduced in (1) may be regarded as a special example of functional autoregression (of order 1), which is often denoted by FAR(1). From this view, our time series (w_t) of demeaned state densities is just a special functional autoregressive process (of order 1). We may define more general FAR(p) models and processes in a similar way. FAR has the same structure as, and may thus be regarded as, a generalization of the vector autoregression (VAR) that has been widely used in time series econometric modeling. In our FAR model, we simply have functional variables, which are generally infinite dimensional. This contrasts to VAR, where we have a finite number of variables included in the model. FAR and VAR are the same in many aspects such as motivation and intended interpretation. However, technically it is more difficult and involved to deal with FAR than with VAR. For instance, the infinite dimensionality of FAR introduces the ill-posed inverse problem, which makes it harder to do inference. Interested readers are referred to Bosq (2000) and the references therein for more detailed exposition on FAR.

³Of course, we may consider the model $f_t = \mu + Af_{t-1} + \varepsilon_t$ given directly in terms of the state densities (f_t) themselves instead of the demeaned state densities (w_t) , where μ is an additional unknown parameter function corresponding to a constant term in the usual regression. In this paper, we use a more compact presentation in (1) to focus on the estimation and interpretation of the operator A .

⁴Here and elsewhere in this paper, we use the notation 1 to denote the constant function taking value 1, instead of the identity function.

One advantage of our FAR model for state densities introduced in (1) is that it may be used to investigate the intertemporal dynamics of the moments of state distributions represented by their densities. To explain our approach, let $v \in H$, and consider a coordinate-process version of our model given as

$$\langle v, w_t \rangle = \langle v, Aw_{t-1} \rangle + \langle v, \varepsilon_t \rangle = \langle A^*v, w_{t-1} \rangle + \varepsilon_t(v), \quad (2)$$

where A^* is the adjoint of A and $(\varepsilon_t(v))$ is a scalar white noise process. The joint operator of a linear operator A on a Hilbert space H is the linear operator A^* on H such that for any $z, w \in H$, $\langle z, Aw \rangle = \langle A^*z, w \rangle$. For any $v \in H$ given, we may interpret A^*v as the response function of $\langle v, w_t \rangle$ to an impulse to w_{t-1} given by a Dirac delta function. Note that $\langle v, w_t \rangle$ would increase by $\langle A^*v, \delta_x \rangle = (A^*v)(x)$, in response to an impulse to w_{t-1} given by the Dirac delta function δ_x with a spike at x .

We may use (2) to analyze the intertemporal dynamics of various moments of state distributions.⁵ For ι_p defined as $\iota_p(x) = x^p$ for $p = 1, 2, \dots$, we have

$$\langle \iota_p, w_t \rangle = \int_C x^p w_t(x) dx,$$

which is the p -th moment of the demeaned state distribution at time t . Furthermore, if we write $A^*\iota_p = \sum_{q=1}^{\infty} c_{p,q} \iota_q$ for a given p with some real sequence $(c_{p,q})$ for $q = 1, 2, \dots$, then it follows that

$$\langle A^*\iota_p, w_{t-1} \rangle = \sum_{q=1}^{\infty} c_{p,q} \int_C x^q w_{t-1}(x) dx,$$

which is an infinite linear combination of all integral moments of the lagged demeaned state distribution represented by (w_{t-1}) . Consequently, it follows from (2) that

$$\int_C x^p w_t(x) dx = \sum_{q=1}^{\infty} c_{p,q} \int_C x^q w_{t-1}(x) dx + \varepsilon_t(\iota_p),$$

which can be used to analyze the moment dynamics of state distributions.

Two special cases appear to be worth mentioning. In what follows, we let (\mathcal{F}_t) be a filtration, to which (w_t) is adapted. If we assume $A^*\iota_2 = \alpha\iota_2$ for some constant $0 < \alpha < 1$, it follows that $\langle \iota_2, w_t \rangle = \alpha\langle \iota_2, w_{t-1} \rangle + \varepsilon_t(\iota_2)$, from which we may deduce that

$$\mathbb{E}(\langle \iota_2, w_t \rangle | \mathcal{F}_{t-1}) = \alpha\langle \iota_2, w_{t-1} \rangle.$$

⁵Here by convention we extend the inner product in H to define $\langle v, w_t \rangle$ for $v \in L^2(C)$. This convention will be made throughout the paper.

Therefore, (2) with $v = \iota_2$ reduces essentially to an ARCH model. If in addition to $A^*\iota_2 = \alpha\iota_2$, we let $A^*\iota_1 = \alpha_1\iota_1 + \alpha_2\iota_2$, then we have

$$\langle \iota_1, w_t \rangle = \beta_1 \langle \iota_1, w_{t-1} \rangle + \beta_2 \mathbb{E}(\langle \iota_2, w_t \rangle | \mathcal{F}_{t-1}) + \varepsilon_t(\iota_1)$$

where $\beta_1 = \alpha_1$ and $\beta_2 = \alpha_2/\alpha$. In this case, (2) with $v = \iota_1$ becomes an ARCH-M model studied in Engle et al. (1987). Clearly, our model provides a much more general framework within which we analyze various moment dynamics of general state distributions.

In addition, we may have

$$v(x) = 1_B(x),$$

where 1_B is an indicator function for a subset B of C . In this case, $\langle v, w_t \rangle = \int_B w_t(x) dx$ in (2) represents the probability of the underlying state being at B , in terms of the deviation from its expected value. More specifically, for instance, the probability of the underlying state being at a left tail can be analyzed if we set $B = (-\infty, -\tau) \cap C$ with $\tau > 0$. We can also choose $B = ((-\infty, -\tau) \cup (\tau, \infty)) \cap C$, so that the probability of the underlying state being at an extreme value can be studied.

Throughout this paper we make the following assumptions:

Assumption 1. *We assume that*

- (a) *A is a compact linear operator such that $\|A^k\| < 1$ for some $k \geq 1$, and,*
- (b) *(ε_t) are i.i.d. such that $\mathbb{E}\varepsilon_t = 0$ and $\mathbb{E}\|\varepsilon_t\|^4 < \infty$, and are independent of w_0 .*

We use $\|\cdot\|$ to denote the operator norm for bounded operators on H . The operator norm of A on H is defined as $\|A\| = \sup_{v \in H} \|Av\| / \|v\|$. Recall that a linear operator is compact if it maps the open unit ball to a set that has a compact closure. Since a compact operator on a Hilbert space can be represented as the limit (in operator norm) of a sequence of finite dimensional operators, many features of the finite dimensional matrix theory naturally generalize to compact linear operators. As a consequence, we may regard A essentially as an infinite dimensional matrix. As shown in Bosq (2000), $\|A^k\| < 1$ for some $k \geq 1$ if and only if $\|A^k\| \leq ab^k$ with some $a > 0$ and $0 < b < 1$ for all $k \geq 0$. This condition implies that FAR(1) introduced in (1) has a unique stationary solution for (w_t) . The i.i.d. assumption for (ε_t) introduced above is more restrictive than is necessary and can be relaxed to less stringent conditions. Many of our subsequent results hold for sequences that are only serially uncorrelated.

2.2. Some Preliminaries

For an H -valued random variable w , we define its mean by the element $\mathbb{E}w \in H$ such that we have, for all $v \in H$,

$$\langle v, \mathbb{E}w \rangle = \mathbb{E}\langle v, w \rangle. \quad (3)$$

If $\mathbb{E}\|w\| < \infty$, then $\mathbb{E}w$ exists and is unique. It can be shown that many of the properties of the usual expectation hold in the functional random variable case. For example, the expectation operator is linear. In particular, if A is a bounded linear operator in H , we have that $\mathbb{E}(Aw) = A\mathbb{E}w$. In addition, we have that $\|\mathbb{E}w\| \leq \mathbb{E}\|w\|$. We define the covariance operator of two H -valued random variables z and w by the linear operator $\mathbb{E}(z \otimes w)$ on $H \times H$, for which we have

$$[\mathbb{E}(z \otimes w)]v = \mathbb{E}\langle w, v \rangle z$$

for all $v \in H$. Note that we have utilized the tensor product notation defined by $(u \otimes v)(\cdot) = \langle v, \cdot \rangle u$ for any $u, v \in H$. If $\mathbb{E}\|z\|^2 < \infty$ and $\mathbb{E}\|w\|^2 < \infty$, then $\mathbb{E}(z \otimes w)$ exists and is unique. Naturally, we may call $\mathbb{E}(w \otimes w)$ the variance operator of w .

An H -valued random variable w is called Gaussian if the random variable $\langle v, w \rangle$ is Gaussian for all $v \in H$. We denote by $\mathbb{N}(\mu, \Sigma)$ a Gaussian random variable with mean μ and variance operator Σ .

An operator A on a separable Hilbert space H is compact if and only if it can be written as

$$A = \sum_{k=1}^{\infty} \kappa_k (u_k \otimes v_k) \quad (4)$$

for some orthonormal systems (u_k) and (v_k) of H and a sequence (κ_k) of nonnegative numbers tending to 0. The compact linear operator is called nuclear if $\sum_{k=1}^{\infty} |\kappa_k| < \infty$, and Hilbert-Schmidt if $\sum_{k=1}^{\infty} \kappa_k^2 < \infty$. It is well known that $\mathbb{E}(z \otimes w)$ is nuclear, and therefore compact. Furthermore, the variance operator $\mathbb{E}(w \otimes w)$ is self-adjoint and nonnegative (strictly positive if non-degenerate), and admits the spectral representation

$$\mathbb{E}(w \otimes w) = \sum_{k=1}^{\infty} \lambda_k (v_k \otimes v_k) \quad (5)$$

for some nonnegative (strictly positive if non-degenerate) sequence (λ_k) and some orthonormal basis (v_k) of H .

The spectral representation of the autoregressive operator as in (4) gives rise to interesting interpretations of the autoregressive operator. Note that, under (4), our model in (1)

can be written as

$$w_t = \sum_{k=1}^{\infty} \kappa_k \langle v_k, w_{t-1} \rangle u_k + \varepsilon_t.$$

If we order (κ_k) and correspondingly (v_k) and (u_k) such that $\kappa_1 > \kappa_2 > \dots$, then v_1 may be viewed as the direction in which w_{t-1} mainly affects w_t . The corresponding feature of the distribution generated from v_1 will be called the *leading progressive feature*. On the other hand, u_1 represents the direction in which w_t is mainly affected by w_{t-1} . The corresponding feature of the distribution generated from u_1 will be called the *leading regressive feature*. For example, if $v_1 = \iota_2$ and $u_1 = \iota_1$ so that the leading progressive feature is the second moment and the leading regressive feature is the first moment, then in this process the second moment of w_{t-1} mainly affects the distribution of w_t and the first moment of w_t is mainly affected by the distribution of w_{t-1} .

A sequence (ε_t) of H -valued random variables is called a white noise if $\mathbb{E}(\varepsilon_t \otimes \varepsilon_t)$ is the same for all t and $\mathbb{E}(\varepsilon_t \otimes \varepsilon_s) = 0$ for all $t \neq s$. It is easy to see that if (ε_t) is a white noise, then for any $v \in H$, the scalar process $(\varepsilon_t(v))$, $\varepsilon_t(v) = \langle v, \varepsilon_t \rangle$, is a white noise. Under Assumption 1, the model (1) has a unique stationary solution for (w_t) , which is given by

$$w_t = \sum_{k=0}^{\infty} A^k \varepsilon_{t-k},$$

where (ε_t) is a white noise process. See, for instance, Bosq (2000) for more details. It is therefore convenient to define the autocovariance function of (w_t) by

$$\Gamma(k) = \mathbb{E}(w_t \otimes w_{t-k})$$

for $k \in \mathbb{Z}$, as in the analysis of finite dimensional time series. Of particular interest are $P = \Gamma(1)$ and $Q = \Gamma(0)$. It is easy to deduce that $\Gamma(k) = A^k Q$ for $k \geq 1$ and $\Gamma(k) = \Gamma(-k)^*$ for all k . In particular, we have

$$P = A Q, \tag{6}$$

which will be used to estimate A .

It is well known that compact linear operators on infinite dimensional Hilbert spaces are not invertible. Therefore one may not directly use the relationship in (6) to define the autoregressive operator A as $A = P Q^{-1}$ since Q^{-1} is not well defined on H . In fact, if the kernel of Q is $\{0\}$, then Q^{-1} is well defined only on $\mathcal{R}(Q) = \{u \in H \mid \sum_{k=1}^{\infty} \langle u, v_k \rangle^2 / \lambda_k^2 < \infty\}$, which is a proper subspace of H . Consequently, we have $A = P Q^{-1}$ well defined only on the restricted domain $\mathcal{D} = \mathcal{R}(Q)$. This problem is often referred to as the ill-posed inverse

problem.

The standard method to circumvent this problem in the functional data analysis literature is to restrict the definition of A in a finite dimensional subspace of H . To explain this method, let $\lambda_1 > \lambda_2 > \dots > 0$ and define V_K to be the subspace of H spanned by the K eigenvectors v_1, \dots, v_K associated with the eigenvalues $\lambda_1, \dots, \lambda_K$ of Q . Let $Q_K = \Pi_K Q \Pi_K$ where Π_K denotes the projection onto V_K and define

$$Q_K^+ = \sum_{k=1}^K \lambda_k^{-1} (v_k \otimes v_k), \quad (7)$$

i.e., the inverse of Q on V_K .⁶

Now let

$$A_K = P Q_K^+, \quad (8)$$

which is the autoregressive operator A restricted to the subspace V_K of H . Note that V_K is generated by the first K principal components of (w_t) . Since (λ_k) decreases to zero, we may expect that A_K approximates A well if the dimension of V_K increases. The estimator of A , which will be introduced in the next section, is indeed the sample analogue estimator of A_K in (8), and we let K increase as T increases.

3. Estimation

In most applications, (f_t) are not directly observed and should therefore be estimated before we look at the FAR model specified by (1). We suppose that N observations from the probability density f_t are available so that we may estimate f_t consistently for each $t = 1, 2, \dots$. In what follows, we denote by \hat{f}_t the consistent estimator of f_t based on N observations and define $\Delta_t = \hat{f}_t - f_t$ for $t = 1, \dots, T$. As before, we let (λ_k) be the ordered eigenvalues of Q .

Assumption 2. *We assume that*

- (a) $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots > 0$,
- (b) $\|f_t\|, \|\Delta_t\| \leq M$ a.s. for some constant $M > 0$, and,
- (c) $\sup_{t \geq 1} \mathbb{E} \|\Delta_t\|^2 = O(N^{-r})$ for some $r > 0$.

Assumption 2 is very mild and appears to be widely satisfied in practical applications. The condition in part (a) holds if and only if the kernel of Q contains only the origin, which is met as long as the distribution of (f_t) is non-degenerate. Boundedness of (f_t) in

⁶See, e.g., Benatia et al. (2017) for the use of other methods of regularization.

part (b) is expected to hold in many practical applications, since (f_t) is a family of density functions. This is not strictly necessary, and is only required for an optimal rate in the strong laws of large numbers we establish in Section 3.1 and 3.2. For unbounded (f_t) , we have a reduced rate of convergence for the strong laws of large numbers. However, the L^2 -convergence results in Section 3.1 and 3.2 hold for unbounded (f_t) without any modification. The condition in part (c) holds in many applications. For instance, if (f_t) are estimated for each t by the kernel density estimator from N observations, we may expect for all t that $\mathbb{E} \|\Delta_t\|^2 = O(N^{-r})$ for some $0 < r < 2/5$ under very general regularity conditions which allow in particular for dependency among N observations. See, e.g., Bosq (1998, Theorem 2.2). One may expect even higher rates of convergence if higher-order kernels are used. See, e.g., Wand and Jones (1995, Section 2.8) .

3.1. Mean and Autocovariance Operators

First we estimate $\mathbb{E}f$ by the sample average of (f_t) , i.e., by

$$\bar{f} = \frac{1}{T} \sum_{t=1}^T \hat{f}_t.$$

The following result holds.

Lemma 1. *Let Assumptions 1 and 2 hold. If $N \geq cT^{1/r}$ for some constant $c > 0$, then we have*

$$\mathbb{E} \|\bar{f} - \mathbb{E}f\|^2 = O(T^{-1})$$

as $T \rightarrow \infty$. Furthermore, if $N > cT^{2/r} \ln^s T$ for some constants $c > 0$ and $s > 0$, then we have

$$\|\bar{f} - \mathbb{E}f\| = O\left(T^{-1/2} \ln^{1/2} T\right) \quad a.s.$$

as $T \rightarrow \infty$.

Lemma 1 establishes the L^2 and a.s. consistency for the sample mean \bar{f} . It shows in particular that using estimated densities (\hat{f}_t) in place of (f_t) does not affect the convergence rates established in Theorem 3.7 and Corollary 3.2 of Bosq (2000) as long as the number of observations used to estimate (f_t) is sufficiently large. The conditions in Assumptions 1 and 2 are sufficient to yield optimal rates in the laws of large numbers here.

Similarly, we may use the sample analogue to consistently estimate the autocovariance

operators Q and P . Let $\hat{w}_t = \hat{f}_t - \bar{f}$ for $t = 1, 2, \dots$, and define

$$\hat{Q} = \frac{1}{T} \sum_{t=1}^T (\hat{w}_t \otimes \hat{w}_t)$$

and

$$\hat{P} = \frac{1}{T} \sum_{t=1}^T (\hat{w}_t \otimes \hat{w}_{t-1}).$$

Then we have

Theorem 2. *Let Assumptions 1 and 2 hold. If $N \geq cT^{1/r}$ for some constant c , then*

$$\mathbb{E} \left\| \hat{Q} - Q \right\|^2 = O(T^{-1}),$$

and

$$\mathbb{E} \left\| \hat{P} - P \right\|^2 = O(T^{-1})$$

as $T \rightarrow \infty$. Moreover, if $N > cT^{2/r} \ln^s T$ for some constants $c > 0$ and $s > 0$, then

$$\left\| \hat{Q} - Q \right\| = O\left(T^{-1/2} \ln^{1/2} T\right) \quad a.s.$$

and

$$\left\| \hat{P} - P \right\| = O\left(T^{-1/2} \ln^{1/2} T\right) \quad a.s.$$

as $T \rightarrow \infty$.

Theorem 2 shows that the sample analogue estimators \hat{Q} and \hat{P} based on the estimated density functions (\hat{f}_t) are consistent, and have the same rates of convergence as those based on the true density functions (f_t), which have been established in Theorem 4.1 and Corollary 4.1 of Bosq (2000).

3.2. Eigenvalues and Eigenvectors

The implementation of our methodology requires the estimation of eigenvalues and eigenvectors of Q , and Q can be consistently estimated by \hat{Q} by Theorem 2. Naturally, the eigenvalues and eigenvectors of Q are estimated by those of \hat{Q} , which we denote by $(\hat{\lambda}_k, \hat{v}_k)$. We assume that the estimated eigenvalues are distinct and order them so that $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots$ ⁷.

⁷We could potentially allow for multiplicity. However, in that case the eigenvectors could not be uniquely identified even after normalization. As a consequence we need to show consistency of eigenspaces, which complicates the notations and proofs substantially. In this paper, we assume for convenience that the eigenvalues are distinct.

The eigenvalues and eigenvectors of \hat{Q} are the pairs $(\hat{\lambda}, \hat{v})$ such that $\hat{Q}\hat{v} = \hat{\lambda}\hat{v}$.

We may solve the eigen-problem either by discretizing (\hat{w}_t) or by expanding (\hat{w}_t) into linear combinations of an orthonormal basis of H and obtaining the matrix representation of \hat{Q} with respect to that basis. Either way, we transform the original problem into the problem of computing eigenvalues and eigenvectors of a matrix. Readers are referred to Section 8.4 in Ramsey and Silverman (2005) for more details.

For definiteness, let us assume that the eigenspace corresponding to (λ_k) is one dimensional for each k from now on. This implies that Assumption 2(a) should be replaced by $\lambda_1 > \lambda_2 > \dots > 0$. We define $v'_k = \text{sgn}\langle \hat{v}_k, v_k \rangle v_k$. Note that since both (v_k) and $(-v_k)$ are eigenvectors corresponding to λ_k , the introduction of (v'_k) is essential for definiteness of eigenvectors. The following lemma from Bosq (2000) associates $|\hat{\lambda}_k - \lambda_k|$ and $\|\hat{v}_k - v'_k\|$ with $\|\hat{Q} - Q\|$.

Lemma 3. *Let Assumptions 1 and 2(a) hold. We have*

$$\sup_{k \geq 1} |\hat{\lambda}_k - \lambda_k| \leq \|\hat{Q} - Q\|$$

and

$$\|\hat{v}_k - v'_k\| \leq \tau_k \|\hat{Q} - Q\|,$$

where $\tau_1 = 2\sqrt{2}(\lambda_1 - \lambda_2)^{-1}$, and $\tau_k = 2\sqrt{2} \max\{(\lambda_{k-1} - \lambda_k)^{-1}, (\lambda_k - \lambda_{k+1})^{-1}\}$ for $k \geq 2$.

Let $\tau(K) = \sup_{1 \leq k \leq K} (\lambda_k - \lambda_{k+1})^{-1}$ and set K to be a function of T , i.e., K_T such that $K_T \rightarrow \infty$ as $T \rightarrow \infty$. It follows directly from Theorem 2 and Lemma 3 that

Theorem 4. *Let Assumptions 1 and 2 hold. We have*

$$\sup_{k \geq 1} |\hat{\lambda}_k - \lambda_k| = O\left(T^{-1/2} \ln^{1/2} T\right) \quad a.s.$$

as $T \rightarrow \infty$.

Our results in Theorem 4 shows that the convergence rates of the estimated eigenvalues, in the presence of estimation errors for the densities, are the same as in Theorem 4.4 and Corollary 4.3 in Bosq (2000).

3.3. Autoregression and Error Variance Operators

Using the estimators introduced in Sections 3.1 and 3.2, we may consistently estimate the autoregressive operator A by an estimator of A_K given earlier in (8). To be explicit, we

define

$$\hat{Q}_K^+ = \sum_{k=1}^K \hat{\lambda}_k^{-1} (\hat{v}_k \otimes \hat{v}_k)$$

and subsequently

$$\hat{A}_K = \hat{P} \hat{Q}_K^+. \quad (9)$$

Theorem 5. *Let Assumptions 1 and 2 hold. Moreover, suppose $N \geq cT^{2/r} \ln^s T$ for some constants $c > 0$ and $s > 0$, A is Hilbert-Schmidt, and $\frac{\ln T (\sum_{k=1}^K \tau_k)^2}{T \lambda_K^2} \rightarrow 0$ as $T \rightarrow \infty$. Then*

$$\left\| \hat{A}_K - A \right\| \rightarrow 0 \quad a.s.$$

as $T \rightarrow \infty$.

Recall that K is a function of the sample size T such that $K \rightarrow \infty$ as $T \rightarrow \infty$. Theorem 5 establishes the consistency of \hat{A}_K . Consistency as in Theorem 8.8 in Bosq (2000) continues to hold in the presence of estimation errors for the densities.

Now we define

$$\hat{\Sigma} = \frac{1}{T} \sum_{k=1}^T (\hat{\varepsilon}_k \otimes \hat{\varepsilon}_k), \quad (10)$$

where $(\hat{\varepsilon}_k)$ obtained by

$$\hat{\varepsilon}_k = \hat{w}_k - \hat{A}_K \hat{w}_{k-1} \quad (11)$$

are the fitted residuals from the FAR in (1) with (w_t) replaced by (\hat{w}_t) . As is expected from Theorem 5, we have

Corollary 6. *Let the conditions in Theorem 5 hold. Then we have*

$$\left\| \hat{\Sigma} - \Sigma \right\| \rightarrow 0 \quad a.s.$$

as $T \rightarrow \infty$.

Corollary 6 establishes the strong consistency of $\hat{\Sigma}$.

4. Analysis of Distributional Autoregression

This section provides some tools and asymptotics to analyze the distributional autoregression. In particular, we discuss two topics including the forecast of state density and the moment dynamics of state distributions.

4.1. Forecast of State Density

Our model can be used to obtain the forecasts of future state densities. For one-step forecast, we use

$$\hat{w}_{T+1} = \hat{A}_K \hat{w}_T, \quad (12)$$

where \hat{A}_K is the estimator of the autoregressive operator A introduced in (9). Multiple step forecasts can be obtained by recursively applying \hat{A}_K to the forecasted values in the previous steps. Below we develop the asymptotic theory for our one-step forecast.

The development of our asymptotics require some extra technical conditions in addition to what we have in Assumptions 1 and 2.

Assumption 3. *We assume that*

- (a) λ_k is convex in k for large k ,
- (b) $\mathbb{E}\langle v_k, w_t \rangle^4 \leq M \lambda_k^2, k \geq 1$, for some constant $M > 0$ independent of k , and,

Condition (a) is weak and satisfied by many sequences including $\lambda_k = c/k^a$, $\lambda_k = c/(k^a \ln^b k)$ and $\lambda_k = ce^{-ak}$, where a, b and c are some positive constants. Let $\delta_k = \lambda_k - \lambda_{k+1}$. Then δ_k is a decreasing sequence by convexity of λ_k . The condition (b) is met whenever the tail probability of $\langle v_k, w_t \rangle$ decreases fast enough. For instance, if w_t is Gaussian, the condition holds with $M = 3$. Note that $\langle v_k, w_t \rangle$ has mean zero and variance λ_k for $k = 1, 2, \dots$. It may not hold if the distribution of $\langle v_k, w_t \rangle$ for some k has a thicker tail.

In the development of our subsequent theory, we denote by Π_K the projection onto the subspace spanned by the first K principal eigenvectors v_1, \dots, v_K of Q , and denote by $\hat{\Pi}_k$ the projection onto the subspace spanned by the first K principal eigenvectors $\hat{v}_1, \dots, \hat{v}_K$ of \hat{Q} . The asymptotic theory that we develop for the one-step forecast is based on the foundational results due to Mas (2007). To avoid technical difficulties, we follow Mas (2007) to compute the estimator \hat{A}_K using samples only up to time $T - 1$.

Lemma 7. *Let Assumptions 1, 2 and 3 hold. If (a) $N \geq cT^{2/r} \ln^s T$ for some constants $c > 0$ and $s > 0$, (b) $\frac{1}{\sqrt{T}} K^{5/2} \ln^2 K \rightarrow 0$, and (c) $\|Q^{-1/2}A\| < \infty$, then*

$$\sqrt{T/K} \left(\hat{A}_K \hat{w}_T - A \hat{\Pi}_K w_T \right) \rightarrow_d \mathbb{N}(0, \Sigma)$$

as $T \rightarrow \infty$, where $\mathbb{N}(0, \Sigma)$ is a Gaussian random element taking values in the Hilbert space H , with mean zero and variance operator Σ .

The condition (b) requires that K should not increase too fast relative to T . Loosely put, condition (c) requires that A should be at least as smooth as $Q^{1/2}$. If indeed they

have common eigenvectors, then this condition holds if and only if the ordered sequence of eigenvalues of A decays at least as fast as the ordered sequence of eigenvalues of $Q^{1/2}$.

The result in Lemma 7 is not directly applicable to obtain the asymptotic confidence interval for the forecast of (w_t) . In particular, it has a random bias term $(A\hat{\Pi}_K - A)w_T$. This can be remedied, at the expense of a set of additional conditions as in the following theorem.

Theorem 8. *Let the assumptions in Lemma 7 hold. Suppose $\sup_{k \geq 1} \mathbb{E}\langle v_k, \Delta_t \rangle^4 / \lambda_k^2 = O(N^{-2r})$, $(T/K) \sum_{k=K+1}^{\infty} \lambda_k \rightarrow 0$, $\sum_{i=1}^K \sum_{j=K+1}^{\infty} \lambda_i \lambda_j / (\lambda_i - \lambda_j)^2 = o(K)$, and that there exists constants M and $|b| < 1$ such that $\sup_{v_p, v_m} \text{Corr}(\langle w_t, v_p \rangle \langle \Delta_t, v_m \rangle, \langle w_s, v_p \rangle \langle \Delta_s, v_m \rangle) \leq Mb^{|s-t|}$. Then*

$$\sqrt{T/K} \left(\hat{A}_K \hat{w}_T - Aw_T \right) \rightarrow_d \mathbb{N}(0, \Sigma)$$

as $T \rightarrow \infty$.

The newly introduced conditions in Theorem 8 can be restrictive, and depend upon the decaying rate of the eigenvalues of the variance operator Q , which is unobserved, and asymptotic independence of the process $(w_t \otimes \Delta_t)$. However, they are not prohibitively stringent. For instance, the conditions on the eigenvalues are satisfied for $\lambda_k = ce^{-k}$ where c is some positive constant, and it seems that in many applications the eigenvalues indeed exhibit an exponential decay. The conditions on the correlations are satisfied if $(w_t \otimes \Delta_t)$ is a mixingale or ρ -mixing, with geometrically decaying mixing rates.

The asymptotic result in Theorem 8 can be used to construct the asymptotic confidence interval for the one-step forecast of (w_t) . Indeed, we may easily deduce from the result in Theorem 8 that

$$\hat{w}_{T+1} - w_{T+1} = \hat{A}_K \hat{w}_T - Aw_T - \varepsilon_{T+1} \approx_d \mathbb{N} \left(0, \left(1 + \frac{K}{T} \right) \Sigma \right).$$

It follows that for any $v \in H$ the interval forecast for $\langle v, w_{T+1} \rangle$ with confidence level α is given by

$$\left[\langle v, \hat{w}_{T+1} \rangle - z_{\alpha/2} \sqrt{(1 + K/T) \langle v, \Sigma v \rangle}, \langle v, \hat{w}_{T+1} \rangle + z_{\alpha/2} \sqrt{(1 + K/T) \langle v, \Sigma v \rangle} \right]$$

with $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$, where Φ is the cumulative distribution function of the standard normal. Here Σ could be replaced by its consistent estimate in implementation.

4.2. Moment Dynamics of State Distributions

In this section, we demonstrate how we may analyze the moment dynamics of state distributions. The analysis proves useful in learning the dependence structure in the moments of state distributions across time. For the analysis, we define a normalized moment basis (u_k) in H such that u_k is the k -th order polynomial for $k = 1, 2, \dots$, and that

- (a) $\langle 1, u_k \rangle = 0$ for all k , and
- (b) $\langle u_p, Qu_q \rangle = 1$ if $p = q$, and 0 if $p \neq q$, for all p and q .

Such a basis could be obtained by the Gram-Schmidt orthogonalization procedure. For $C = [a, b]$, we may easily see that u_1 is given by $u_1(x) = C_1[x - (a + b)/2]$, where C_1 is a constant determined by $\langle u_1, Qu_1 \rangle = 1$. Moreover, once u_1, u_2, \dots, u_p are obtained, we may readily find u_{p+1} such that $\langle 1, u_{p+1} \rangle = 0$, $\langle u_q, Qu_{p+1} \rangle = 0$ for $q = 1, \dots, p$ and $\langle u_{p+1}, Qu_{p+1} \rangle = 1$. Note that (u_k) is orthonormal with respect to the inner product $\langle \cdot, Q \cdot \rangle$ and generates H . That is, it is an orthonormal basis of H . Since u_k is a polynomial of order k , we shall view it as the function that generates the normalized k -th order moment of state distributions.

Now we consider (2), which defines the dynamics for the v -moment of the underlying state distributions. For any $v \in H$ given, write

$$A^*v = \sum_{k=1}^{\infty} \langle A^*v, Qu_k \rangle u_k = \sum_{k=1}^{\infty} \langle v, AQu_k \rangle u_k,$$

so that we may deduce

$$\langle A^*v, w_{t-1} \rangle = \sum_{k=1}^{\infty} \langle v, AQu_k \rangle \langle u_k, w_{t-1} \rangle, \quad \mathbb{E} \langle A^*v, w_{t-1} \rangle^2 = \sum_{k=1}^{\infty} \langle v, AQu_k \rangle^2, \quad (13)$$

which are quite useful to analyze the moment dynamics of state distributions. Note that we have

$$\mathbb{E} \langle u_p, w_t \rangle \langle u_q, w_t \rangle = \langle u_p, Qu_q \rangle = \begin{cases} 1, & \text{if } p = q, \\ 0, & \text{if } p \neq q \end{cases}$$

by construction. Consequently, for any $v \in H$ given, we have

$$\mathbb{E} \langle v, w_t \rangle^2 = \mathbb{E} \langle A^*v, w_{t-1} \rangle^2 + \mathbb{E} \langle v, \varepsilon_t \rangle^2 = \sum_{k=1}^{\infty} \langle v, AQu_k \rangle^2 + \mathbb{E} \langle v, \varepsilon_t \rangle^2 \quad (14)$$

due to (2) and (13).

Table 1. Forecast Error Measures

| Deviation | Definition |
|-----------|---|
| D_2 | $(\int(\hat{f}(x) - f(x))^2 dx)^{1/2}$ |
| D_1 | $\int \hat{f}(x) - f(x) dx$ |
| D_{ks} | $\sup_x \hat{F}(x) - F(x) $ |
| D_{cm} | $\int(\hat{F}(x) - F(x))^2 dF(x)$ |
| D_m | $ \int x \hat{f}(x) dx - \int x f(x) dx $ |
| D_v | $ \int [x - \int \hat{f}(x) dx]^2 \hat{f}(x) dx - \int [x - \int f(x) dx]^2 f(x) dx $ |

In our analysis of the moment dynamics of state distributions, for any given $v \in H$ we define

$$R_v^2 = 1 - \frac{\mathbb{E}\varepsilon_t^2(v)}{\mathbb{E}\langle v, w_t \rangle^2} = 1 - \frac{\langle v, \Sigma v \rangle}{\langle v, Qv \rangle} \quad (15)$$

to be the R^2 of the v -moment of state distribution, and interpret the ratio

$$\pi_v(k) = \frac{\mathbb{E}\langle v, AQu_k \rangle^2 \langle u_k, w_{t-1} \rangle^2}{\mathbb{E}\langle v, w_t \rangle^2} = \frac{\langle v, AQu_k \rangle^2}{\langle v, Qv \rangle} \quad (16)$$

as the proportion of the variance of the v -moment of the current state distribution contributed by the k -th moment of the past state distribution. Clearly, R_v^2 and $\pi_v(k)$ defined respectively in (15) and (16) can be consistently estimated by

$$\hat{R}_v^2 = 1 - \frac{\langle v, \hat{\Sigma}v \rangle}{\langle v, \hat{Q}v \rangle} \quad \text{and} \quad \hat{\pi}_v(k) = \frac{\langle v, \hat{A}_K \hat{Q}u_k \rangle^2}{\langle v, \hat{Q}v \rangle}$$

for any $v \in H$, where $\hat{\Sigma}$, \hat{Q} and \hat{A}_K are the estimates of Σ , Q and A , respectively, which have been introduced in Section 3.

5. Empirical Applications

In this section we use our methodology to analyze two financial markets. In the first, we study the intra-month distributions of the GBP/USD exchange rate 15-minute log returns. In the second, we study the cross-sectional distributions of monthly returns of stocks listed on the NYSE. We explore various aspects of the distributional dynamics in the foreign exchange and stock markets that would not be revealed if we only look at their aggregate time series.

We also present the forecast performances of our model. For each forecast period, we obtain the predicted density \hat{f} and calculate how much it deviates from the actually

observed density f , using the six measures given in Table 1. \hat{F} and F in the table are the distribution functions corresponding to \hat{f} and f , respectively. The first two measures, namely the L^2 - and the L^1 -deviations, are based on the density functions. The middle two measures, namely the Kolmogorov-Smirnov and the Cramér-von Mises deviations, are based on the distribution functions. The last two measures, namely the (absolute) deviation in mean and in variance, are based on the first two moments of the densities. Notice that all these quantities are always nonnegative, and are zero when $\hat{f} = f$.

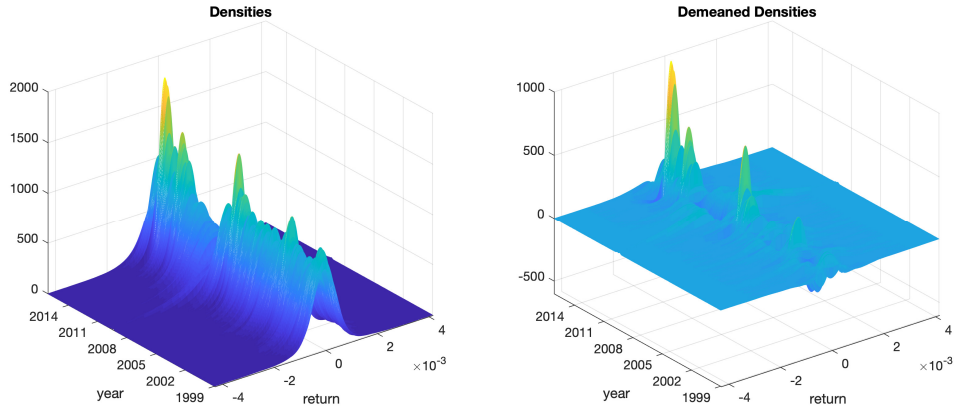
5.1. *Intra-Month Distributions of GBP/USD Exchange Rate 15-Minute Log Returns*

For a high frequency intra-month financial return series, we may consider the individual intra-month returns as draws from a common distribution, which varies by month, under the piecewise stationarity assumption. These draws may be dependent, hence accommodating the case in which the intra-month returns are not independent but are strictly stationary. This view not only provides a justification for applying our model to time-varying return distributions, but also reflects the fact that the intra-month returns exhibit statistical characteristics that do not carry over to longer horizons.

In this application, we look at the intra-month distributions of the GBP/USD exchange rate 15-minute log returns. Since the foreign exchange market operates globally around-the-clock on weekdays, it is more natural to treat every four weeks as a month than to follow the calendar months. We thus set a month in this application to be a four week period. We use data from January 4th, 1999 to April 3rd, 2015, and split the data into a total of 212 months using the four-week convention. The number of observations in each period ranges from 1550 to 1904, with a mean of 1880. We obtain the kernel density estimate for each period using the Epanechnikov kernel with the optimal bandwidth $h_t = 2.3449\hat{\sigma}_t n_t^{-1/5}$, where $\hat{\sigma}_t$ and n_t are respectively the sample standard deviation and the sample size in period t . The support of the densities is set to be $[-0.0043, 0.0043]$, which includes more than 99.9% of the actual observations. We plot the time series of both the densities and the demeaned densities in Figure 1, from which we could easily see the time-varying nature of the corresponding distributions.

To implement our method, we represent the densities with the Daubechies wavelets using 1037 basis functions. Each density is then represented as a 1037-dimensional vector whose coordinates are the wavelet coefficients of the density function. Thus we transform the functional time series into a time series of high dimensional vectors. We obtain the matrix representations of the sample variance operator \hat{Q} and the sample first-order autocovariance operator \hat{P} with respect to that basis, which are respectively the sample covariance and the sample first-order autocovariance matrices of the vector time series. We then estimate the

Fig. 1. Intra-Month GBP/USD Exchange Rate Return Distributions



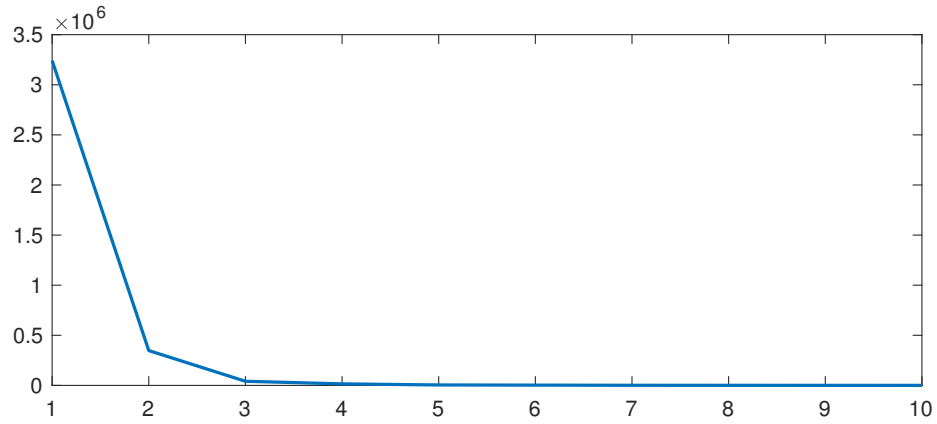
Notes: The left and right panels respectively present the densities and demeaned densities of the intra-month distributions obtained from the GBP/USD exchange rate 15-minute log returns.

eigenvalues and eigenvectors of the variance operator Q respectively by the eigenvalues and eigenvectors of the sample covariance matrix of the vector time series. When we approximate the inverse of the variance operator, we set $K = 4$ to obtain the best rolling out-of-sample forecast performance (see below). This is the data-driven method we propose for choosing the value of the nuisance parameter K . As another justification for our choice of K from the perspective of functional principal component analysis, we note that the first 4 principal components together explain 99.7% of the total variation in the density process. See Figure 2 for the scree plot of the ten largest eigenvalues of the variance operator.

As an illustration of how we may use the tools developed in this paper to perform interesting analyses, we first obtain the leading progressive and regressive features for this density process. Each feature is represented by its values at 1024 points evenly spaced on the support of the distributions. See Figure 3. The leading progressive feature is a concentrating feature: If it is loaded with a positive scale, the distribution will be more concentrated around its center. It also shows that the relative frequency of observations with small returns is most important for determining distributions in the future. The leading regressive feature shows it is also the relative frequency of observations with small returns that is most affected by distributions in the past. Both features indicate that the observations with extreme returns are not important in the distributional dynamics.

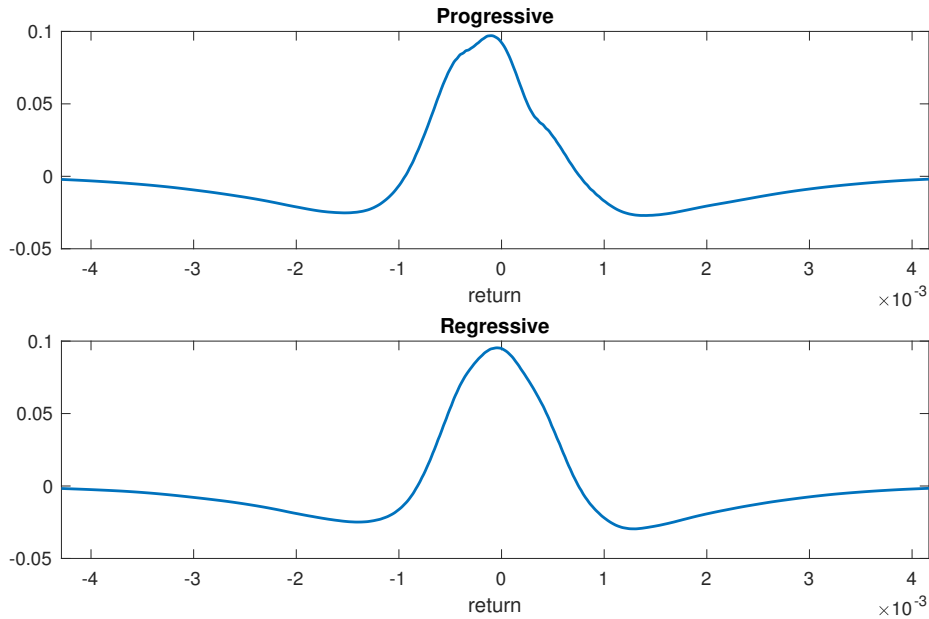
We also illustrate how various aspects in the current distribution respond to shocks to the past distribution. The left two panels in Figure 4 present the current first two moments' responses with respect to Dirac- δ impulses to the previous distribution at different

Fig. 2. Scree Plot for Intra-Month GBP/USD Exchange Rate Return Distributions



Notes: The ten largest eigenvalues of the sample variance operator for the intra-month densities of the GBP/USD exchange rate 15-minute log returns are presented in descending order.

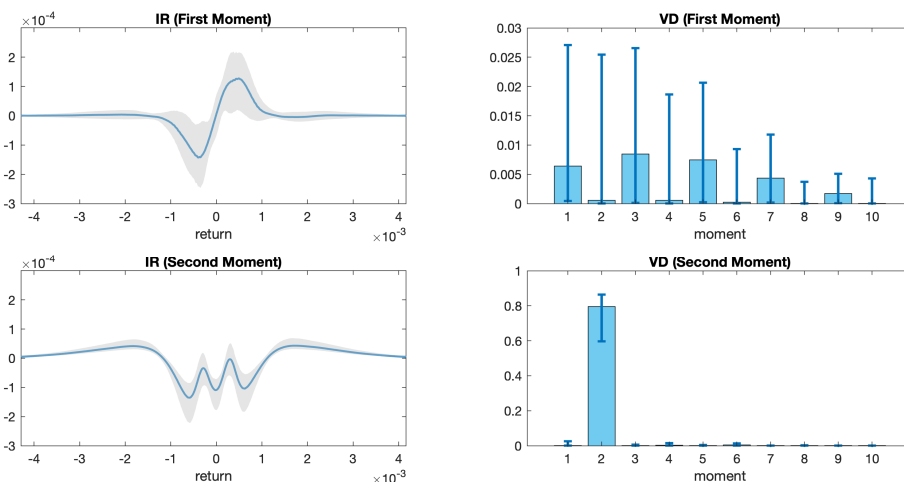
Fig. 3. Leading Features of Intra-Month GBP/USD Exchange Rate Return Distributions



Notes: The upper and lower panels present respectively the leading progressive and regressive features of the intra-month distributions of the GBP/USD exchange rate 15-minute log returns.

levels of return. The shaded areas give the corresponding 95% residual bootstrap confidence bands based on 2000 repetitions. The top left panel shows that the mean of the current distribution is most affected by shocks to the relative frequencies of the observations with small-to-moderate returns in the previous period, and shocks to the relative frequencies of the observations with large or extreme returns in the previous period are not important.

Fig. 4. Impulse Responses and Variance Decompositions for Moments of Intra-Month GBP/USD Exchange Rate Return Distributions

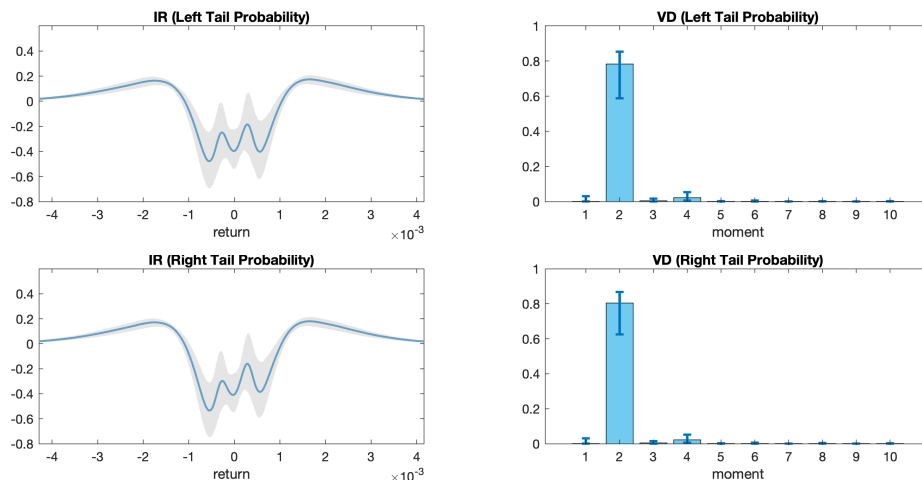


Notes: The left two panels respectively plot the impulse responses of the first moment (upper panel) and the second moment (lower panel) of the current distribution of the GBP/USD exchange rate 15-minute log returns to Dirac- δ impulses to the last month's density function at different levels of return. The shaded areas give the corresponding 95% residual bootstrap confidence bands based on 2000 repetitions. The right two panels depict respectively the proportions of variances of the first moment (upper panel) and the second moment (lower panel) of the current distribution of the GBP/USD exchange rate 15-minute log returns that are explained by the variances of the first ten moments of the last month's distribution, with their 95% residual bootstrap confidence intervals.

Moreover, we see the momentum effect, i.e., a larger number of observations with small-to-moderate negative (positive) returns in the previous period is likely to result in a lower (higher) first moment in the current period. The bottom left panel shows that the second moment of the current distribution, on the other hand, is affected by shocks to the relative frequencies of returns at all levels in the previous period. One may expect a smaller second moment in the current period if there were more observations with small-to-moderate returns in the previous period, and a larger second moment in the current period if there were more observations with large returns in the previous period. The effects of small-to-moderate returns in the previous period on the second moment in the current period are not monotonic. The stabilizing effects of some moderate returns are at least as strong as those of the small returns.

The right two panels give the proportions of the variances of the current first and second moments that are explained by the variances of the first ten moments of the previous distribution, with their 95% residual bootstrap confidence intervals. The corresponding R_v^2 are respectively 0.0380 and 0.8091, with 95% bootstrap confidence intervals [0.0085, 0.1186]

Fig. 5. Impulse Responses and Variance Decompositions for Tail Probabilities of Intra-Month GBP/USD Exchange Rate Return Distributions



Notes: The left two panels respectively plot the impulse responses of the left tail probability (upper panel) and the right tail probability (lower panel) of the current distribution of the GBP/USD exchange rate 15-minute log returns to Dirac- δ impulses to the last month's density function at different levels of return. The shaded areas give the corresponding 95% residual bootstrap confidence bands based on 2000 repetitions. The right two panels depict respectively the proportions of variances of the left tail probability (upper panel) and the right tail probability (lower panel) of the current distribution of the GBP/USD exchange rate 15-minute log returns that are explained by the variances of the first ten moments of the last month's distribution, with their 95% residual bootstrap confidence intervals.

and $[0.6214, 0.8727]$. Though many integral moments, especially those of odd orders, affect the current mean, their overall effect is almost negligible. In contrast, the past second moment, whose variance explains about 80% of the total variance, is very informative about the current second moment, while the other past moments provide almost no information. Our findings here are entirely consistent with the widespread and repeated observations by many empirical researchers that in financial markets the mean is usually difficult to predict, while the volatility process is much more persistent and therefore much more predictable.

Figure 5 gives the impulse responses and the variance decompositions of the tail probabilities. The left tail includes the returns that are smaller than the 5th percentile of the mean density, and the right tail includes the returns that are greater than the 95th percentile of the mean density. In this application, the responses of the two tail probabilities to shocks in the past density are quite symmetric. Shocks to, and subsequent changes in, the relative frequencies of returns at all levels in the previous period matter. A larger number of observations with small-to-moderate returns in the last period tends to presage thinner tails of the current distribution, while a larger number of observations with large returns

Table 2. Forecast Errors for Intra-Month GBP/USD Exchange Rate Return Distributions

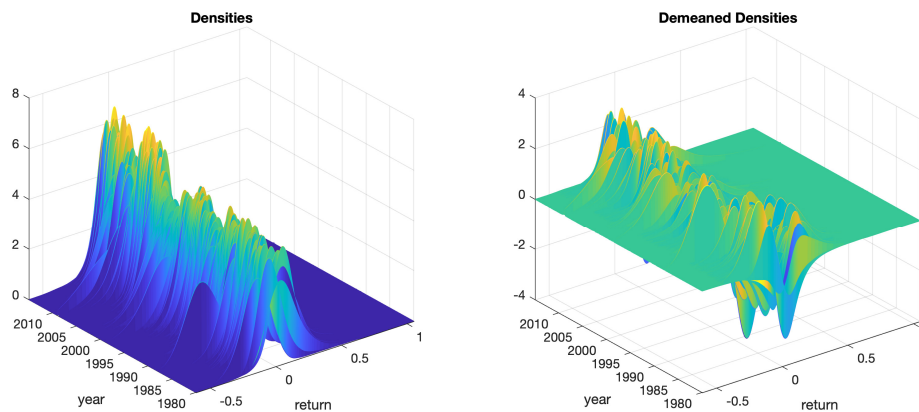
| | FAR | AVE | LAST |
|-------------------|------------|------------|------------|
| D_2 | 2.67(2.15) | 5.54(4.19) | 2.67(2.12) |
| $D_1(10^{-1})$ | 1.14(0.97) | 2.49(2.04) | 1.15(0.90) |
| $D_{ks}(10^{-2})$ | 3.10(2.94) | 6.47(5.53) | 3.28(2.73) |
| $D_{cm}(10^{-3})$ | 0.65(0.35) | 2.75(1.26) | 0.68(0.27) |
| $D_m(10^{-5})$ | 0.82(0.73) | 0.81(0.71) | 1.27(1.14) |
| $D_v(10^{-7})$ | 0.60(0.40) | 1.78(1.87) | 0.58(0.51) |

in the previous period tends to presage thicker tails of the current distribution. The tail responses of small-to-moderate returns are non-monotonic, and are essentially the same as their responses to the second moment. In the variance decomposition analysis, about 80% of the variances of the current tail probabilities are explained by the variances of the past second moment, which shows that the past second moment is very informative about the current tail probabilities.

Finally, to evaluate the prediction performance of our model in this application, we make rolling out-of-sample forecasts of the state densities for the last 50 periods. For each forecast period, we use the data prior to that period to estimate the autoregressive coefficient. We dynamically choose K by cross-validation, using the five periods before the forecast period as the validation periods. We then make one-period-ahead forecasts of the demeaned density as in (12). We add the mean density to the forecasted demeaned density to obtain the density forecast. We normalize the forecasted density by setting its negative values to zero and then rescaling it so that it integrates to one. In the end, we calculate the forecast error. For comparison, we consider two other predictors, which are respectively the time average of all the estimated densities and the estimated density of the last period. We label our functional autoregressive predictor as the FAR predictor, and the other two respectively as the AVE predictor and the LAST predictor. The AVE and the LAST predictors are used as benchmarks: If the true process of the random densities is indeed independent and identically distributed over time, one would expect the AVE predictor to perform well. If the true process is a nonstationary martingale process, one would expect the LAST predictor to perform the best. Comparisons with the two benchmarks reveal information on how far away the actual process is from the i.i.d. scenario and from the nonstationarity scenario.

Table 2 reports the means and medians (in parentheses) of the 50 periods of forecast errors for each of the three predictors. In terms of the first four measures, the FAR predictor performs roughly the same as the LAST predictor, and they outperform the AVE estimator significantly. This suggests that the true process does not resemble an i.i.d. process, and

Fig. 6. Cross-Sectional Distributions of NYSE Stocks Monthly Returns



Notes: The left and right panels respectively present the densities and demeaned densities of the cross-sectional distributions obtained from monthly returns of stocks listed on the New York Stock Exchange.

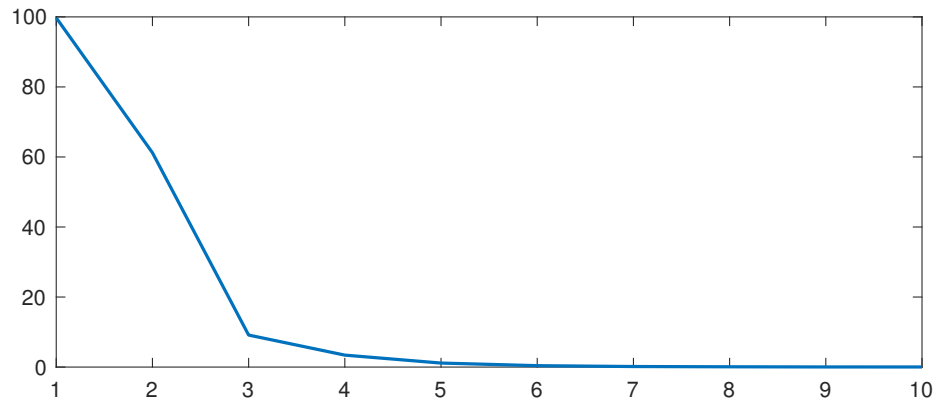
that there is likely to be strong persistency in the density process. In terms of predicting the mean, the FAR predictor performs about the same as the AVE predictor, and much better than the LAST predictor. In terms of predicting the variance, the FAR predictor performs about the same as the LAST predictor, and much better than the AVE estimator. These results corroborate the findings that in financial markets the mean is usually difficult to predict, while the volatility is quite persistent, and therefore much more predictable.

5.2. Cross-Sectional Distributions of NYSE Stocks Monthly Returns

In this application, we look at the cross-sectional distributions of monthly returns of stocks listed on the New York Stock Exchange. In each month, the individual stock returns are viewed as draws from a common distribution. This distribution changes from month to month and is assumed to follow our FAR model (1). We use data from January of 1980 to December of 2014, a total of 420 months. In each month, we only take into account the stocks that are traded. The number of observations each month ranges from 1926 to 3076, with a mean of 2464. We use the same method as in the previous application to estimate the densities. The support of the density in this application is set to be $[-0.6071, 1.0548]$, which includes more than 99.9% of the actual observations. Figure 6 plots the times series of the densities and the demeaned densities of the NYSE stocks monthly returns.

We use the same procedures as in the previous application for model estimation. We set $K = 3$ based on the out-of-sample forecast performance. Figure 7 gives the scree plot of the eigenvalues of the sample variance operator in this application. The first 3

Fig. 7. Scree Plot for Cross-Sectional Distributions of NYSE Stocks Monthly Returns

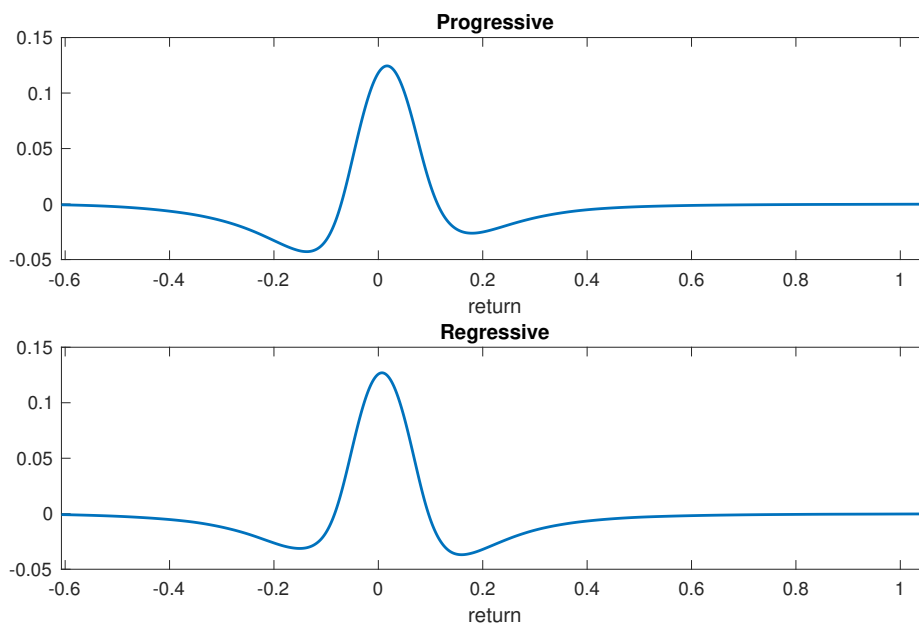


Notes: The ten largest eigenvalues of the sample variance operator for the cross-sectional densities of the NYSE stocks monthly returns are presented in descending order.

principal components explain 97.0% of the total variation in the density process. Figure 8 demonstrates the leading progressive and regressive features for the density process of the NYSE stocks monthly returns. As in the GBP/USD exchange rate application, it is the relative frequency of small returns that is most critical in the distribution dynamics. The small returns play a central role in affecting the future distribution, and they are most affected by the distribution in the previous period.

In Figure 9, we provide the impulse responses and the variance decompositions of the first two moments in this application. The corresponding R_v^2 are 0.0368 and 0.3376 respectively with 95% residual bootstrap confidence intervals [0.0108, 0.0815] and [0.2199, 0.4709], which implies that the first moment is much less predictable than the second moment through the integral moments of the past distribution. Unlike the exchange rate returns application, in which the current mean responds only to changes in the relative frequencies of small returns in the previous period, the current mean in this application responds more significantly to changes in the relative frequencies of moderate-to-large returns in the previous period. The typical momentum effect is generated by stocks with moderately large returns, both positive and negative. For moderately small negative returns, we even find some evidence of a reverse momentum effect, a pattern which is also found in international stock returns by Fama and French (2012). As expected, the second moment of current return distribution decreases as we have more small return stocks in the past distribution. Interestingly, the current second moment responds positively to changes in the relative frequencies of moderate-to-large negative returns, but not to changes in the relative frequencies of positive returns, in the previous period. This is presumably due to the leverage effect, and our result here suggests that the leverage effect is asymmetric, being effective only for negative return

Fig. 8. Leading Features of Cross-Sectional Distributions of NYSE Stocks Monthly Returns

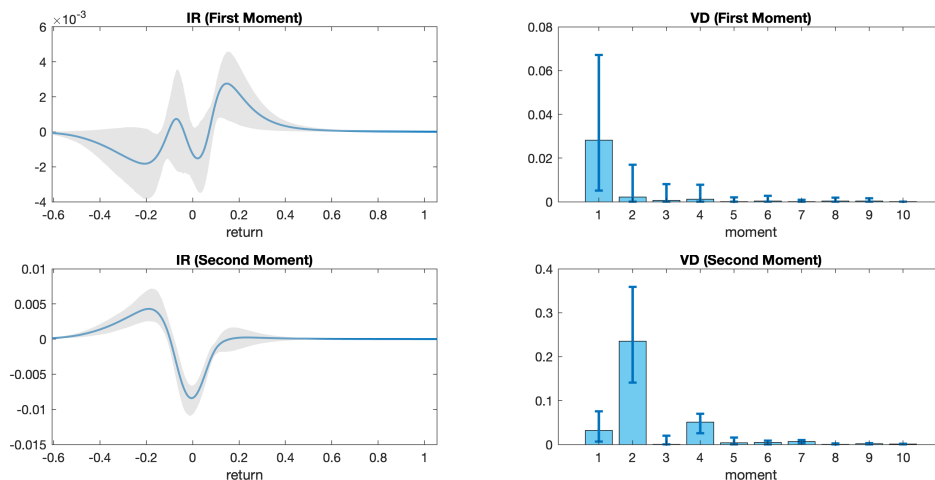


Notes: The upper and lower panels respectively present the leading progressive and regressive features of the cross-sectional distributions of the NYSE stocks monthly returns.

shocks. The past integral moments are not informative about the current mean. They are relatively more informative about the current second moment. However, the relevance is smaller than in the previous application, and the past fourth and first moments appear to play some non-negligible roles.

Figure 10 presents the impulse responses and the variance decompositions of the tail probabilities in this application. The tail events are defined in the same way as in the last application. In contrast with the previous application, the left tail probabilities and the right tail probabilities in this application respond differently to Dirac- δ impulses to the last period's density. As is expected, positive shocks to the relative frequencies of small returns decrease both the left and the right tail probabilities in the next period. Positive shocks to the relative frequencies of moderate-to-large negative returns tend to presage a thicker left tail, whereas positive shocks to the relative frequencies of moderate-to-large positive returns tend to presage a thicker right tail. Of the two effects, the former appears to be much more significant than the latter. Moreover, positive shocks to the relative frequencies of moderate-to-large positive returns decrease the left tail probability, whereas positive shocks to the relative frequencies of moderate-to-large negative returns also increase the right tail probability, though less substantially. On the other hand, the current left tail probability is related to both the past first and second moments, while the current right

Fig. 9. Impulse Responses and Variance Decompositions for Moments of Cross-Sectional Distributions of NYSE Stocks Monthly Returns



Notes: The left two panels respectively plot the impulse responses of the first moment (upper panel) and the second moment (lower panel) of the current distribution of the NYSE stocks monthly returns to Dirac- δ impulses to the last month's density function at different levels of return. The shaded areas give the corresponding 95% residual bootstrap confidence bands based on 2000 repetitions. The right two panels depict respectively the proportions of variances in the first moment (upper panel) and in the second moment (lower panel) of the current distribution of the NYSE stocks monthly returns that are explained by the variances of the first ten moments of the last month's distribution, with their 95% residual bootstrap confidence intervals.

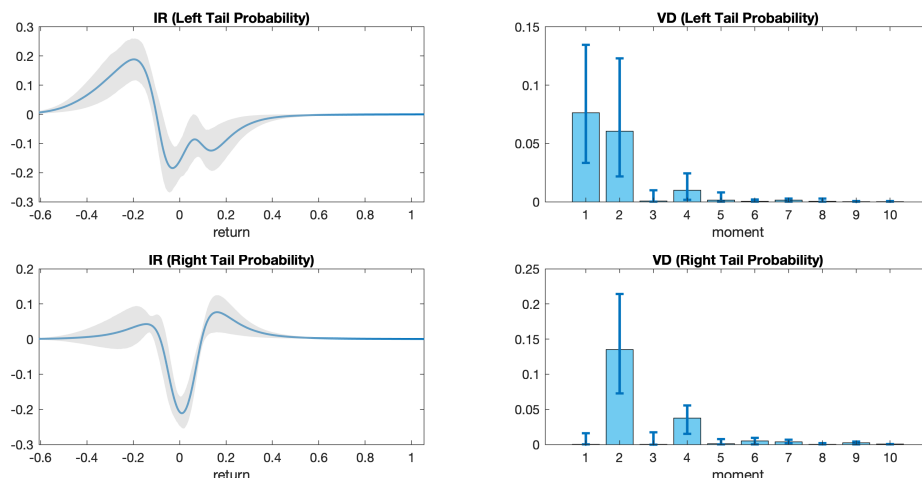
tail probability hinges more on the past second and fourth moments.

In this application, we make rolling out-of-sample forecasts for the last 100 periods since we have longer time series available. Table 3 summarizes the prediction results. In terms of the first four measures, our FAR predictor performs the best while the LAST estimator performs the worst. In terms of predicting the mean, the FAR predictor performs slightly worse than the AVE predictor, and much better than the LAST predictor. In terms of predicting the variance, the FAR predictor performs worse than the LAST predictor, but much better than the AVE predictor. These results suggest that this density process is likely to be stationary, and our model has good forecast power. However, there may be some particular directions in which the process is close to nonstationary, as volatility appears to be quite persistent.

6. Simulations

In this section we study the finite sample properties of our FAR estimator by conducting out-of-sample forecast using simulated data. To make our simulations more realistic and

Fig. 10. Impulse Responses and Variance Decompositions for Tail Probabilities of Cross-Sectional Distributions of NYSE Stocks Monthly Returns



Notes: The left two panels respectively plot the impulse responses of the left tail probability (upper panel) and the right tail probability (lower panel) of the current distribution of the NYSE stocks monthly returns to Dirac- δ impulses to the last month's density function at different levels of return. The shaded areas give the corresponding 95% residual bootstrap confidence bands based on 2000 repetitions. The right two panels depict respectively the proportions of variances in the left tail probability (upper panel) and in the right tail probability (lower panel) of the current distribution of the NYSE stocks monthly returns that are explained by the variances of the first ten moments of the last month's distribution, with their 95% residual bootstrap confidence intervals.

Table 3. Forecast Errors for Cross-Sectional Distributions of NYSE Stocks Monthly Returns

| | FAR | AVE | LAST |
|-------------------|------------|------------|------------|
| $D_2(10^{-1})$ | 5.17(4.79) | 5.95(5.95) | 6.55(5.16) |
| $D_1(10^{-1})$ | 3.44(2.92) | 3.93(3.73) | 4.29(3.57) |
| $D_{ks}(10^{-1})$ | 1.62(1.42) | 1.75(1.46) | 2.10(1.79) |
| $D_{cm}(10^{-2})$ | 1.91(1.03) | 1.94(0.94) | 3.31(1.64) |
| $D_m(10^{-2})$ | 4.05(2.73) | 3.90(2.59) | 5.25(3.73) |
| $D_v(10^{-3})$ | 4.04(3.30) | 7.10(6.47) | 2.89(1.79) |

practically relevant, we use the estimation results from the previous two empirical applications and generate data that mimics the estimated empirical models as closely as possible.

To be specific, for each application, we take the estimated coefficient \hat{A}_K to be the autoregressive coefficient in the FAR(1) data generating process given in (1) for the process (w_t) . In each iteration of simulations, the error process is bootstrapped from the demeaned residuals of the estimated FAR(1) model, given in (11). Once we obtain the simulated demeaned densities (w_t) , we add back the time average of the estimated densities \bar{f} and

normalize them as in the forecast procedures in the two applications to obtain simulated density functions. In each simulation, we start from $\varepsilon_0 = 0$, simulate 1000 periods and take the last $T + 1$ periods. The actual values for T will be specified below. We take the time series of the last $T + 1$ density functions as the true process of densities. We then simulate N observations from the density in each period using acceptance (or rejection) sampling. The actual values for N will be specified below. Acceptance sampling is a widely used sampling method, especially in Bayesian statistics, when one can not directly sample from the distribution but the corresponding density function is known. The researcher samples from a proposed distribution which is different from the target distribution, and accepts or rejects each draw with a probability that is determined by the ratio of the proposal and target density functions evaluated at the draw. In our sampling procedure, we use the uniform distribution on the support as the proposal distribution. We take the accepted draws as the simulated observations.

We implement our functional autoregression using the simulated observations from the first T periods, leaving the last period for forecast performance evaluation. The estimation procedure as well as the parameter settings are the same as in the empirical applications, except that in the kernel density estimation step, we use the Normal kernel instead of the Epanechnikov kernel, with the corresponding optimal bandwidth. The Normal kernel performs better in our simulations in that it gives smaller kernel density estimation error. Once we obtain the estimated autoregressive coefficient, we do one-step-ahead forecast for the last period and compare our forecast with the true density in that period. The forecast procedure is exactly the same as in the two applications in Section 5. We evaluate the difference of our forecast from the true density using the six criteria introduced in the previous section.

We set T to be 50, 100, 200, or 500, and N to be 100, 200, 500 or 5000. For each pair of (T, N) , the mean and median of the forecast errors based on 2000 iterations of simulations are reported in Table 4 and Table 5. For convenience, we label the simulations based on the two estimated models as the “FOREX simulations” and the “NYSE simulations” respectively. We make the following observations.

First, for both sets of simulations, the forecast performance is not very sensitive to the number of periods (T) we have in the estimation. Once we have at least 100 periods, results become highly stable. It seems that a total of 50 periods is still acceptable, although the forecast errors are slightly bigger.

Second, forecast performances are more sensitive to the number of observations (N) in each period. For both sets of simulations and for all values of T , when we increase the number of observations in each period, the forecast errors decrease non-negligibly. This suggests

that a large part of the forecast error comes from the density estimation procedure. This agrees with the usual observation that kernel density estimation converges at a relatively slow rate.

Third, for both sets of simulations, and for all combinations of T and N , the FAR estimator performs best in terms of the first four criteria. This implies that our method has good curve fitting property, in terms of both the density function (the first two criteria) and the distribution function (the middle two criteria). In the FOREX simulations, the FAR predictor performs roughly the same as the AVE predictor, and much better than the LAST estimator in predicting the mean. In predicting the variance, the FAR predictor performs roughly the same as the LAST predictor, and much better than the AVE predictor. In the NYSE simulations, the FAR predictor has the best performance in predicting both the mean and the variance.

To summarize, we find that the FAR predictor forecasts densities accurately, and that the forecast is stable, even with a relatively short time series.

7. Conclusions

This paper presents a model for time-varying densities that follow an autoregressive stationary process in some function space. We explore the properties and implications of this model, showing how it relates to many familiar time series models such as the ARCH-type models. Our approach is highly nonparametric, in that it does not assume any parametric form for the underlying distributions. We have outlined the procedures to estimate the model assuming that the densities are not observable, and have shown the consistency of the estimator as both the sample size for the density estimation and the time horizon go to infinity. Furthermore, we illustrate how to forecast densities using our model, and establish the asymptotic normality of the predictor. We develop tools such as progressive/regressive analysis, impulse response analysis and variance decomposition to help study the dynamics in various aspects of the distribution process. We discover several interesting features in the distribution process of the intra-month GBP/USD exchange rate 15-minute log returns and in the distribution process of the NYSE stocks monthly returns. Finally, Monte Carlo simulations show that the FAR density predictor outperforms its competitors. In sum, the FAR modeling for time series of state distributions is a very general framework that is highly flexible, readily applicable, and it provides a powerful method to extract information of a distributional process and generate accurate forecasts.

Table 4. Monte Carlo Simulations on Out-of-Sample Forecasts (GBP/USD Exchange Rate)

| | | $N = 100$ | | | $N = 200$ | | | $N = 500$ | | | $N = 5000$ | | |
|-----------|-------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| | | FAR | AVE | LAST | FAR | AVE | LAST | FAR | AVE | LAST | FAR | AVE | LAST |
| $T = 50$ | D_2 | 3.37 (3.00) | 4.46(3.87) | 4.10(3.85) | 3.03 (2.57) | 4.46(3.81) | 3.47(3.11) | 2.62 (2.21) | 4.27(3.62) | 2.89(2.52) | 2.12 (1.74) | 4.13(3.54) | 2.15(1.73) |
| | $D_1(10^{-1})$ | 1.63 (1.47) | 2.15(1.89) | 2.02(1.92) | 1.44 (1.26) | 2.14(1.88) | 1.70(1.56) | 1.25 (1.08) | 2.06(1.80) | 1.40(1.24) | 1.01 (0.84) | 1.99(1.75) | 1.03(0.85) |
| | $D_{ks}(10^{-2})$ | 4.48 (4.03) | 5.65(4.98) | 6.27(5.94) | 3.90 (3.43) | 5.56(4.91) | 5.08(4.62) | 3.37 (2.89) | 5.36(4.69) | 4.07(3.70) | 2.79 (2.41) | 5.19(4.49) | 2.92(2.50) |
| | $D_{cm}(10^{-3})$ | 0.98 (0.57) | 1.88(0.98) | 1.79(1.30) | 0.76 (0.41) | 1.85(0.96) | 1.17(0.78) | 0.58 (0.30) | 1.72(0.88) | 0.76(0.48) | 0.42 (0.20) | 1.62(0.82) | 0.43(0.22) |
| | $D_m(10^{-5})$ | 1.98(1.61) | 1.23 (1.03) | 5.31(4.12) | 1.61(1.31) | 1.10 (0.91) | 3.96(3.19) | 1.24(0.99) | 1.04 (0.87) | 2.71(2.15) | 1.06(0.86) | 1.01 (0.84) | 1.48(1.16) |
| | $D_v(10^{-7})$ | 1.38(1.11) | 2.07(1.78) | 1.24 (0.85) | 1.17(0.89) | 2.03(1.77) | 1.04 (0.69) | 1.07(0.85) | 2.03(1.77) | 0.90 (0.62) | 0.82(0.58) | 1.99(1.71) | 0.76 (0.50) |
| $T = 100$ | D_2 | 3.10 (2.65) | 4.35(3.72) | 4.12(3.81) | 2.93 (2.51) | 4.47(3.84) | 3.51(3.18) | 2.50 (2.16) | 4.40(3.82) | 2.85(2.47) | 2.21 (1.80) | 4.33(3.63) | 2.30(1.80) |
| | $D_1(10^{-1})$ | 1.50 (1.32) | 2.11(1.84) | 2.05(1.93) | 1.39 (1.21) | 2.15(1.89) | 1.72(1.57) | 1.19 (1.03) | 2.13(1.87) | 1.38(1.24) | 1.04 (0.87) | 2.08(1.83) | 1.09(0.88) |
| | $D_{ks}(10^{-2})$ | 4.04 (3.61) | 5.51(4.86) | 6.33(5.90) | 3.74 (3.30) | 5.60(4.92) | 5.16(4.77) | 3.18 (2.78) | 5.51(4.83) | 4.04(3.64) | 2.88 (2.48) | 5.41(4.75) | 3.08(2.59) |
| | $D_{cm}(10^{-3})$ | 0.82 (0.46) | 1.82(0.92) | 1.83(1.26) | 0.72 (0.37) | 1.87(0.95) | 1.19(0.83) | 0.52 (0.27) | 1.82(0.95) | 0.73(0.48) | 0.46 (0.22) | 1.76(0.91) | 0.50(0.24) |
| | $D_m(10^{-5})$ | 1.54(1.23) | 1.11 (0.92) | 5.42(4.25) | 1.34(1.09) | 1.07 (0.87) | 4.08(3.20) | 1.14(0.96) | 1.02 (0.85) | 2.79(2.21) | 1.01(0.84) | 0.99 (0.83) | 1.48(1.20) |
| | $D_v(10^{-7})$ | 1.31(1.06) | 2.10(1.82) | 1.27 (0.87) | 1.17(0.92) | 2.10(1.86) | 1.08 (0.77) | 1.02(0.78) | 2.13(1.95) | 0.91 (0.61) | 0.79(0.56) | 2.06(1.78) | 0.77 (0.50) |
| $T = 200$ | D_2 | 3.13 (2.69) | 4.53(3.87) | 4.11(3.76) | 2.75 (2.38) | 4.50(3.80) | 3.42(3.11) | 2.45 (2.04) | 4.37(3.72) | 2.86(2.46) | 2.02 (1.56) | 4.20(3.61) | 2.17(1.68) |
| | $D_1(10^{-1})$ | 1.50 (1.31) | 2.19(1.87) | 2.03(1.88) | 1.31 (1.16) | 2.17(1.86) | 1.68(1.54) | 1.16 (0.99) | 2.11(1.83) | 1.38(1.22) | 0.95 (0.76) | 2.03(1.81) | 1.04(0.83) |
| | $D_{ks}(10^{-2})$ | 3.95 (3.47) | 5.67(4.89) | 6.31(5.82) | 3.47 (3.11) | 5.63(4.84) | 5.04(4.63) | 3.11 (2.66) | 5.48(4.76) | 4.04(3.63) | 2.65 (2.17) | 5.27(4.70) | 2.93(2.47) |
| | $D_{cm}(10^{-3})$ | 0.82 (0.42) | 1.95(0.94) | 1.83(1.24) | 0.61 (0.34) | 1.90(0.95) | 1.13(0.76) | 0.51 (0.25) | 1.82(0.93) | 0.74(0.47) | 0.40 (0.17) | 1.66(0.88) | 0.45(0.22) |
| | $D_m(10^{-5})$ | 1.26(1.03) | 1.02 (0.85) | 5.58(4.45) | 1.13(0.94) | 1.02 (0.86) | 3.96(3.14) | 1.03(0.89) | 0.97 (0.83) | 2.74(2.26) | 0.96 (0.78) | 0.96(0.79) | 1.45(1.20) |
| | $D_v(10^{-7})$ | 1.27(1.02) | 2.15(1.90) | 1.27 (0.88) | 1.11(0.90) | 2.15(1.93) | 1.05 (0.72) | 1.00(0.73) | 2.11(1.89) | 0.89 (0.60) | 0.76(0.53) | 2.07(1.87) | 0.75 (0.51) |
| $T = 500$ | D_2 | 3.02 (2.65) | 4.49(3.83) | 4.05(3.78) | 2.77 (2.39) | 4.44(3.82) | 3.46(3.14) | 2.38 (2.03) | 4.36(3.69) | 2.83(2.47) | 2.02 (1.59) | 4.31(3.70) | 2.17(1.73) |
| | $D_1(10^{-1})$ | 1.44 (1.30) | 2.17(1.90) | 2.00(1.86) | 1.31 (1.17) | 2.14(1.86) | 1.69(1.55) | 1.13 (0.97) | 2.10(1.85) | 1.37(1.21) | 0.95 (0.78) | 2.08(1.81) | 1.03(0.84) |
| | $D_{ks}(10^{-2})$ | 3.77 (3.38) | 5.62(4.95) | 6.27(5.83) | 3.46 (3.07) | 5.56(4.85) | 5.11(4.73) | 3.03 (2.59) | 5.47(4.77) | 3.98(3.62) | 2.65 (2.28) | 5.37(4.71) | 2.94(2.55) |
| | $D_{cm}(10^{-3})$ | 0.74 (0.40) | 1.91(0.97) | 1.80(1.27) | 0.63 (0.34) | 1.87(0.92) | 1.16(0.81) | 0.48 (0.24) | 1.80(0.91) | 0.71(0.46) | 0.39 (0.19) | 1.75(0.89) | 0.44(0.23) |
| | $D_m(10^{-5})$ | 1.11(0.95) | 1.01 (0.85) | 5.58(4.49) | 1.05(0.90) | 0.99 (0.85) | 4.07(3.26) | 1.01(0.85) | 1.00 (0.84) | 2.64(2.11) | 0.96(0.80) | 0.96 (0.78) | 1.50(1.21) |
| | $D_v(10^{-7})$ | 1.24(1.00) | 2.13(1.93) | 1.21 (0.83) | 1.12(0.84) | 2.13(1.97) | 1.04 (0.71) | 0.97(0.72) | 2.13(1.98) | 0.89 (0.61) | 0.73 (0.51) | 2.08(1.90) | 0.73(0.49) |

Table 5. Monte Carlo Simulations on Out-of-Sample Forecasts (NYSE Stock Returns)

| | | $N = 100$ | | | $N = 200$ | | | $N = 500$ | | | $N = 5000$ | | |
|-----------|-------------------|--------------------|------------|------------|--------------------|--------------------|------------|--------------------|--------------------|------------|--------------------|------------|------------|
| | | FAR | AVE | LAST | FAR | AVE | LAST | FAR | AVE | LAST | FAR | AVE | LAST |
| $T = 50$ | $D_2(10^{-1})$ | 4.25 (3.97) | 4.40(4.27) | 5.37(4.82) | 4.15 (3.94) | 4.37(4.13) | 5.20(4.74) | 3.98 (3.70) | 4.21(3.99) | 4.99(4.45) | 3.95 (3.70) | 4.29(4.12) | 4.88(4.44) |
| | $D_1(10^{-1})$ | 2.80 (2.63) | 2.90(2.79) | 3.66(3.30) | 2.74 (2.56) | 2.89(2.74) | 3.53(3.21) | 2.63 (2.45) | 2.78(2.65) | 3.35(2.99) | 2.61 (2.44) | 2.85(2.75) | 3.24(2.91) |
| | $D_{ks}(10^{-1})$ | 1.28 (1.19) | 1.31(1.25) | 1.59(1.40) | 1.26 (1.18) | 1.30(1.21) | 1.56(1.38) | 1.22 (1.12) | 1.25(1.17) | 1.52(1.32) | 1.22 (1.12) | 1.28(1.23) | 1.50(1.32) |
| | $D_{cm}(10^{-2})$ | 0.94 (0.60) | 0.95(0.66) | 1.61(0.83) | 0.97(0.62) | 0.96 (0.64) | 1.62(0.83) | 0.92(0.57) | 0.91 (0.59) | 1.58(0.79) | 0.96 (0.58) | 0.99(0.65) | 1.56(0.78) |
| | $D_m(10^{-2})$ | 2.83 (2.62) | 2.83(2.58) | 3.63(3.18) | 2.86(2.64) | 2.81 (2.62) | 3.60(3.10) | 2.78(2.50) | 2.73 (2.50) | 3.51(2.98) | 2.82 (2.51) | 2.83(2.57) | 3.45(2.99) |
| | $D_v(10^{-3})$ | 4.71 (4.07) | 5.33(4.75) | 5.40(4.14) | 4.21 (3.68) | 5.32(4.78) | 4.67(3.72) | 3.93 (3.25) | 5.12(4.56) | 4.27(3.32) | 3.63 (2.97) | 5.12(4.54) | 4.02(3.23) |
| $T = 100$ | $D_2(10^{-1})$ | 4.16 (3.91) | 4.40(4.15) | 5.37(4.91) | 4.08 (3.81) | 4.37(4.15) | 5.10(4.58) | 4.01 (3.79) | 4.37(4.25) | 5.07(4.56) | 3.89 (3.68) | 4.25(4.12) | 4.89(4.42) |
| | $D_1(10^{-1})$ | 2.74 (2.57) | 2.90(2.78) | 3.67(3.34) | 2.68 (2.49) | 2.89(2.75) | 3.45(3.09) | 2.64 (2.52) | 2.89(2.80) | 3.40(3.01) | 2.56 (2.43) | 2.82(2.70) | 3.24(2.91) |
| | $D_{ks}(10^{-1})$ | 1.26 (1.18) | 1.30(1.22) | 1.60(1.41) | 1.24 (1.14) | 1.30(1.22) | 1.52(1.29) | 1.24 (1.18) | 1.31(1.24) | 1.55(1.36) | 1.20 (1.11) | 1.27(1.18) | 1.50(1.33) |
| | $D_{cm}(10^{-2})$ | 0.93 (0.60) | 0.95(0.64) | 1.67(0.83) | 0.92 (0.56) | 0.96(0.63) | 1.52(0.68) | 0.94 (0.63) | 0.98(0.69) | 1.61(0.83) | 0.91 (0.56) | 0.95(0.62) | 1.56(0.82) |
| | $D_m(10^{-2})$ | 2.81 (2.58) | 2.82(2.62) | 3.69(3.09) | 2.78 (2.49) | 2.82(2.56) | 3.48(2.86) | 2.85 (2.67) | 2.87(2.70) | 3.59(3.21) | 2.75 (2.52) | 2.78(2.55) | 3.45(2.97) |
| | $D_v(10^{-3})$ | 4.46 (3.99) | 5.42(4.97) | 5.13(3.90) | 4.21 (3.51) | 5.27(4.59) | 4.84(3.72) | 3.80 (3.11) | 5.16(4.53) | 4.28(3.45) | 3.55 (2.80) | 5.00(4.41) | 4.03(3.21) |
| $T = 200$ | $D_2(10^{-1})$ | 4.13 (3.87) | 4.39(4.23) | 5.30(4.73) | 4.03 (3.72) | 4.35(4.20) | 5.14(4.55) | 4.10 (3.85) | 4.45(4.26) | 5.15(4.61) | 3.83 (3.60) | 4.23(4.10) | 4.93(4.40) |
| | $D_1(10^{-1})$ | 2.70 (2.54) | 2.89(2.78) | 3.61(3.23) | 2.64 (2.45) | 2.87(2.79) | 3.46(3.09) | 2.68 (2.54) | 2.94(2.86) | 3.44(3.06) | 2.53 (2.35) | 2.81(2.73) | 3.26(2.87) |
| | $D_{ks}(10^{-1})$ | 1.25 (1.17) | 1.30(1.22) | 1.56(1.35) | 1.23 (1.13) | 1.29(1.21) | 1.53(1.34) | 1.26 (1.20) | 1.33(1.29) | 1.57(1.39) | 1.19 (1.09) | 1.27(1.18) | 1.51(1.33) |
| | $D_{cm}(10^{-2})$ | 0.90 (0.58) | 0.95(0.63) | 1.56(0.79) | 0.90 (0.56) | 0.95(0.62) | 1.56(0.80) | 0.97 (0.66) | 1.02(0.74) | 1.65(0.88) | 0.89 (0.56) | 0.95(0.64) | 1.57(0.80) |
| | $D_m(10^{-2})$ | 2.77 (2.55) | 2.81(2.54) | 3.54(3.08) | 2.76 (2.48) | 2.78(2.57) | 3.53(3.07) | 2.89 (2.74) | 2.93(2.81) | 3.64(3.24) | 2.75 (2.50) | 2.80(2.58) | 3.49(3.03) |
| | $D_v(10^{-3})$ | 4.50 (3.95) | 5.37(4.77) | 5.30(4.05) | 4.08 (3.41) | 5.27(4.70) | 4.80(3.66) | 3.92 (3.22) | 5.27(4.59) | 4.40(3.36) | 3.44 (2.72) | 4.98(4.38) | 4.00(3.21) |
| $T = 500$ | $D_2(10^{-1})$ | 4.17 (4.00) | 4.46(4.32) | 5.44(4.96) | 4.15 (3.95) | 4.49(4.31) | 5.29(4.80) | 3.90 (3.71) | 4.28(4.11) | 4.96(4.46) | 3.81 (3.50) | 4.20(3.95) | 4.85(4.27) |
| | $D_1(10^{-1})$ | 2.72 (2.62) | 2.94(2.88) | 3.71(3.39) | 2.71 (2.58) | 2.95(2.88) | 3.58(3.21) | 2.57 (2.43) | 2.84(2.73) | 3.33(2.99) | 2.50 (2.29) | 2.78(2.62) | 3.20(2.82) |
| | $D_{ks}(10^{-1})$ | 1.27 (1.20) | 1.33(1.27) | 1.61(1.41) | 1.27 (1.20) | 1.34(1.28) | 1.60(1.42) | 1.20 (1.12) | 1.28(1.21) | 1.51(1.32) | 1.17 (1.06) | 1.25(1.17) | 1.49(1.28) |
| | $D_{cm}(10^{-2})$ | 0.92 (0.63) | 0.97(0.68) | 1.67(0.89) | 0.95 (0.66) | 1.01(0.74) | 1.67(0.88) | 0.90 (0.57) | 0.96(0.64) | 1.54(0.75) | 0.87 (0.52) | 0.94(0.62) | 1.53(0.71) |
| | $D_m(10^{-2})$ | 2.84 (2.64) | 2.87(2.71) | 3.68(3.25) | 2.87 (2.75) | 2.92(2.78) | 3.69(3.32) | 2.76 (2.56) | 2.82(2.54) | 3.48(3.01) | 2.71 (2.53) | 2.77(2.58) | 3.42(2.86) |
| | $D_v(10^{-3})$ | 4.47 (3.81) | 5.50(4.91) | 5.24(3.98) | 4.24 (3.46) | 5.42(4.85) | 4.88(3.73) | 3.81 (3.11) | 5.23(4.65) | 4.33(3.46) | 3.51 (2.82) | 4.90(4.25) | 3.99(3.04) |

Appendix: Mathematical Proofs

Proof of Lemma 1. Let $\tilde{f} = (1/T) \sum_{t=1}^T f_t$. We first show the L^2 consistency. Notice that

$$\|\bar{f} - \mathbb{E}f\|^2 = \|\bar{f} - \tilde{f} + \tilde{f} - \mathbb{E}f\|^2 \leq 2 \left(\|\bar{f} - \tilde{f}\|^2 + \|\tilde{f} - \mathbb{E}f\|^2 \right).$$

By Assumption 2(c), we have that

$$\mathbb{E} \|\bar{f} - \tilde{f}\|^2 = \mathbb{E} \left\| \frac{1}{T} \sum_{i=1}^T (\hat{f}_t - f_t) \right\|^2 \leq \mathbb{E} \frac{1}{T} \sum_{i=1}^T \|\Delta_t\|^2 \leq \sup_t \mathbb{E} \|\Delta_t\|^2 = O(N^{-r}).$$

If $N \geq cT^{1/r}$ for some constant $c > 0$, then we have that $\mathbb{E} \|\bar{f} - \tilde{f}\|^2 = O(T^{-1})$. By Theorem 3.7 in Bosq (2000), we have that $\mathbb{E} \|\tilde{f} - \mathbb{E}f\|^2 = O(T^{-1})$. Therefore, $\mathbb{E} \|\bar{f} - \mathbb{E}f\|^2 = O(T^{-1})$.

Next we show the a.s. consistency. Notice that

$$\|\bar{f} - \mathbb{E}f\| \leq \|\bar{f} - \tilde{f}\| + \|\tilde{f} - \mathbb{E}f\|.$$

By Chebyshev's inequality, for any $\epsilon > 0$,

$$\mathbb{P} \left\{ \left(\frac{T}{\ln T} \right)^{1/2} \|\bar{f} - \tilde{f}\| \geq \epsilon \right\} \leq \frac{T}{\epsilon^2 \ln T} \mathbb{E} \|\bar{f} - \tilde{f}\|^2 = \frac{T}{\epsilon^2 \ln T} O(N^{-r}).$$

If $N > cT^{2/r} \ln^s T$ for some constants $c > 0$ and $s > 0$, then

$$\mathbb{P} \left\{ \left(\frac{T}{\ln T} \right)^{1/2} \|\bar{f} - \tilde{f}\| \geq \epsilon \right\} = O \left(\frac{1}{T \ln^\delta T} \right)$$

where $\delta = 1 + sr > 1$. Since $\sum_{T=2}^{\infty} 1/T \ln^\delta T$ converges for any $\delta > 1$, by the Borel-Cantelli lemma, we have that

$$\mathbb{P} \left\{ \limsup_{T \rightarrow \infty} \left(\frac{T}{\ln T} \right)^{1/2} \|\bar{f} - \tilde{f}\| \geq \epsilon \right\} = 0.$$

This implies that $\|\bar{f} - \tilde{f}\| = o(T^{-1/2} \ln^{1/2} T)$ a.s.. By Corollary 3.2 in Bosq (2000), we have that $\|\tilde{f} - \mathbb{E}f\| = O(T^{-1/2} \ln^{1/2} T)$ a.s.. Therefore, we conclude that $\|\bar{f} - \mathbb{E}f\| = O(T^{-1/2} \ln^{1/2} T)$ a.s.. \square

Proof of Theorem 2. We first prove the L^2 consistency of \hat{Q} . By Assumption 2(b), we have that $\|\mathbb{E}f\| \leq \mathbb{E}\|f_t\| \leq M$. Then for all t ,

$$\|w_t\| = \|f_t - \mathbb{E}f\| \leq \|f_t\| + \|\mathbb{E}f\| \leq 2M \quad \text{a.s.} \quad (17)$$

Similarly, we have that for all t ,

$$\|\hat{f}_t\| = \|f_t + \Delta_t\| \leq 2M \quad \text{a.s.}$$

and therefore

$$\|\hat{w}_t\| = \|\hat{f}_t - \bar{f}\| = \left\| \hat{f}_t - \frac{1}{T} \sum_{i=1}^T \hat{f}_i \right\| \leq 4M \quad \text{a.s.} \quad (18)$$

By Assumption 2(c) and Lemma 1,

$$\mathbb{E}\|\hat{w}_t - w_t\|^2 = \|\hat{f}_t - \bar{f} - f_t + \mathbb{E}f\|^2 \leq 2 \left(\mathbb{E}\|\hat{f}_t - f_t\|^2 + \mathbb{E}\|\bar{f} - \mathbb{E}f\|^2 \right) = O(T^{-1}). \quad (19)$$

Let

$$\tilde{Q} = \frac{1}{T} \sum_{t=1}^T w_t \otimes w_t.$$

then

$$\begin{aligned} \mathbb{E}\|\hat{Q} - \tilde{Q}\|^2 &= \mathbb{E}\left\| \frac{1}{T} \sum_{t=1}^T (\hat{w}_t \otimes \hat{w}_t - w_t \otimes w_t) \right\|^2 \\ &= \frac{1}{T^2} \mathbb{E}\left\| \sum_{t=1}^T [\hat{w}_t \otimes (\hat{w}_t - w_t) + (\hat{w}_t - w_t) \otimes w_t] \right\|^2 \\ &\leq \frac{2}{T} \mathbb{E}\left(\sum_{t=1}^T \|\hat{w}_t \otimes (\hat{w}_t - w_t)\|^2 + \sum_{t=1}^T \|(\hat{w}_t - w_t) \otimes w_t\|^2 \right) \\ &\leq \frac{2}{T} \mathbb{E}\left(\sum_{t=1}^T (\|\hat{w}_t\|^2 + \|w_t\|^2) \|\hat{w}_t - w_t\|^2 \right) \\ &\leq \frac{40M^2}{T} \sum_{t=1}^T \mathbb{E}\|\hat{w}_t - w_t\|^2 = O(T^{-1}). \end{aligned}$$

By Theorem 4.1 in Bosq (2000), we have that

$$\mathbb{E}\|\tilde{Q} - Q\|^2 = O(T^{-1}).$$

Therefore,

$$\mathbb{E} \left\| \hat{Q} - Q \right\|^2 \leq 2 \left(\mathbb{E} \left\| \hat{Q} - \tilde{Q} \right\|^2 + \mathbb{E} \left\| \tilde{Q} - Q \right\|^2 \right) = O(T^{-1})$$

as $T \rightarrow \infty$.

Next we show the a.s. consistency of \hat{Q} . First note that

$$\begin{aligned} \left\| \hat{Q} - \tilde{Q} \right\| &= \left\| \frac{1}{T} \sum_{t=1}^T (\hat{w}_t \otimes \hat{w}_t - w_t \otimes w_t) \right\| \\ &= \left\| \frac{1}{T} \sum_{t=1}^T [\hat{w}_t \otimes (\hat{w}_t - w_t) + (\hat{w}_t - w_t) \otimes w_t] \right\| \\ &\leq \frac{1}{T} \sum_{t=1}^T (\|\hat{w}_t\| + \|w_t\|) \|\hat{w}_t - w_t\| \\ &\leq \frac{6M}{T} \sum_{t=1}^T \|\hat{w}_t - w_t\| \quad \text{a.s..} \end{aligned}$$

Since

$$\mathbb{E} \left(\frac{1}{T} \sum_{t=1}^T \|\hat{f}_t - f_t\| \right)^2 \leq \mathbb{E} \frac{1}{T^2} \cdot T \sum_{t=1}^T \|\Delta_t\|^2 \leq \sup_t \mathbb{E} \|\Delta_t\|^2 = O(N^{-r}),$$

by Chebyshev's inequality, we have that for any $\epsilon > 0$,

$$\mathbb{P} \left\{ \left(\frac{T}{\ln T} \right)^{1/2} \frac{1}{T} \sum_{t=1}^T \|\hat{f}_t - f_t\| \geq \epsilon \right\} \leq \frac{T}{\epsilon^2 \ln T} \mathbb{E} \left(\frac{1}{T} \sum_{t=1}^T \|\hat{f}_t - f_t\| \right)^2 = \frac{T}{\epsilon^2 \ln T} O(N^{-r}).$$

If $N > cT^{2/r} \ln^s T$ for some constants $c > 0$ and $s > 0$, by the Borel-Cantelli lemma, we have that

$$\mathbb{P} \left\{ \limsup_{T \rightarrow \infty} \left(\frac{T}{\ln T} \right)^{1/2} \frac{1}{T} \sum_{t=1}^T \|\hat{f}_t - f_t\| \geq \epsilon \right\} = 0.$$

This implies that

$$\frac{1}{T} \sum_{t=1}^T \|\hat{f}_t - f_t\| = o\left(T^{-1/2} \ln^{1/2} T\right) \quad \text{a.s..}$$

By Lemma 1, $\|\bar{f} - \mathbb{E}f\| = O\left(T^{-1/2} \ln^{1/2} T\right)$ a.s.. Then

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \|\hat{w}_t - w_t\| &= \frac{1}{T} \sum_{t=1}^T \left\| \hat{f}_t - \bar{f} - f_t + \mathbb{E}f \right\| \\ &\leq \frac{1}{T} \sum_{t=1}^T \left\| \hat{f}_t - f_t \right\| + \|\bar{f} - \mathbb{E}f\| \\ &= O\left(T^{-1/2} \ln^{1/2} T\right) \quad \text{a.s..} \end{aligned} \tag{20}$$

By Corollary 4.1 in Bosq (2000), $\|\tilde{Q} - Q\| = O\left(T^{-1/2} \ln^{1/2} T\right)$ a.s.. Therefore

$$\|\hat{Q} - Q\| \leq \|\hat{Q} - \tilde{Q}\| + \|\tilde{Q} - Q\| = O\left(T^{-1/2} \ln^{1/2} T\right) \quad \text{a.s..}$$

The proof for consistency of \hat{P} is similar and therefore is omitted. The proof makes uses of Theorem 4.7 and 4.8 in Bosq (2000). \square

Proof of Lemma 3. See Lemma 4.2 and Lemma 4.3 in Bosq (2000). \square

Proof of Theorem 4. It follows immediately from Theorem 2 and Lemma 3. \square

Proof of Theorem 5. Notice that

$$\begin{aligned} \|\hat{A}_K - A\| &= \left\| \hat{P} \sum_{k=1}^K \hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) - A \right\| \\ &\leq \|D_1\| + \|D_2\| + \|D_3\| + \|D_4\| \end{aligned}$$

where

$$D_1 = \hat{P} \sum_{k=1}^K \hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) - P \sum_{k=1}^K \hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k),$$

$$D_2 = P \sum_{k=1}^K \hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) - P \sum_{k=1}^K \lambda_k^{-1}(\hat{v}_k \otimes \hat{v}_k),$$

$$D_3 = P \sum_{k=1}^K \lambda_k^{-1}(\hat{v}_k \otimes \hat{v}_k) - P \sum_{k=1}^K \lambda_k^{-1}(v_k \otimes v_k),$$

and

$$D_4 = P \sum_{k=1}^K \lambda_k^{-1}(v_k \otimes v_k) - A.$$

Since $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots > 0$, we have that

$$\begin{aligned} \|D_1\| &\leq \left\| \hat{P} - P \right\| \left\| \sum_{i=1}^K \hat{\lambda}_i^{-1} (\hat{v}_i \otimes \hat{v}_i) \right\| \\ &= \hat{\lambda}_K^{-1} \left\| \hat{P} - P \right\| \\ &\leq \lambda_K^{-1} \left\| \hat{P} - P \right\| + \left| \hat{\lambda}_K^{-1} - \lambda_K^{-1} \right| \left\| \hat{P} - P \right\|. \end{aligned}$$

Note that

$$\sum_{k=1}^K \tau_k \geq 2\sqrt{2}(\lambda_K - \lambda_{K+1})^{-1} \geq 2\sqrt{2}\lambda_K^{-1},$$

then by assumption we have that

$$\frac{\ln T \left(\sum_{k=1}^K \tau_k \right)^2}{T\lambda_K^2} \geq \frac{8 \ln T}{T\lambda_K^4} \rightarrow 0$$

as $T \rightarrow \infty$. This implies that

$$\lambda_K^{-1} = o\left(T^{1/4} \ln^{-1/4} T\right). \quad (21)$$

By Theorem 2, we have that $\left\| \hat{P} - P \right\| = O\left(T^{-1/2} \ln^{1/2} T\right)$ a.s.. Then

$$\lambda_K^{-1} \left\| \hat{P} - P \right\| \rightarrow 0 \quad \text{a.s.} \quad (22)$$

as $T \rightarrow \infty$.

Next we notice that

$$\mathbb{P}\left(\liminf_{k \rightarrow \infty} \left\{ \hat{\lambda}_k \geq \frac{\lambda_k}{2} \right\}\right) = 1, \quad (23)$$

for if otherwise, then with positive probability, $\hat{\lambda}_k < \lambda_k/2$ infinitely often. This implies that with positive probability, $\left| \hat{\lambda}_k - \lambda_k \right| > \lambda_k/2$ infinitely often. Then by (21), we have that with positive probability, $\limsup_{T \rightarrow \infty} T^{1/4} \ln^{-1/4} T \left| \hat{\lambda}_k - \lambda_k \right| = \infty$. This contradicts with the conclusion in Theorem 4. Therefore, (23) holds. That is, almost surely, $\hat{\lambda}_K \geq \lambda_K/2$ for K large enough.

Now for such K ,

$$\begin{aligned}
\left| \hat{\lambda}_K^{-1} - \lambda_K^{-1} \right| \left\| \hat{P} - P \right\| &= \frac{|\hat{\lambda}_K - \lambda_K|}{\hat{\lambda}_K \lambda_K} \left\| \hat{P} - P \right\| \\
&\leq 2 \frac{|\hat{\lambda}_K - \lambda_K|}{\lambda_K^2} \left\| \hat{P} - P \right\| \\
&\leq \frac{2 \left\| \hat{Q} - Q \right\| \left\| \hat{P} - P \right\|}{\lambda_K^2}.
\end{aligned}$$

Then by Theorem 2 and equation (21),

$$\left| \hat{\lambda}_K^{-1} - \lambda_K^{-1} \right| \left\| \hat{P} - P \right\| \rightarrow 0 \quad \text{a.s.} \quad (24)$$

Equations (22) and (24) together imply that

$$\|D_1\| \rightarrow 0 \quad \text{a.s.}$$

as $T \rightarrow \infty$.

For K large enough,

$$\begin{aligned}
\|D_2\| &\leq \|P\| \left\| \sum_{k=1}^K \left(\hat{\lambda}_k^{-1} - \lambda_k^{-1} \right) (\hat{v}_k \otimes \hat{v}_k) \right\| \\
&= \max_{1 \leq k \leq K} \left| \hat{\lambda}_k^{-1} - \lambda_k^{-1} \right| \|P\| \\
&\leq \frac{\sup_{k \geq 1} |\hat{\lambda}_k - \lambda_k| \|P\|}{\hat{\lambda}_K \lambda_K} \\
&\leq \frac{2 \left\| \hat{Q} - Q \right\| \|P\|}{\lambda_K^2}.
\end{aligned}$$

By Theorem 2 and equation (21),

$$\|D_2\| \rightarrow 0 \quad \text{a.s.}$$

as $T \rightarrow \infty$.

By Lemma 3,

$$\begin{aligned}
\|D_3\| &\leq \|P\| \sum_{k=1}^K \lambda_k^{-1} \|\hat{v}_k \otimes \hat{v}_k - v'_k \otimes v'_k\| \\
&= \|P\| \sum_{k=1}^K \lambda_k^{-1} \|\hat{v}_k \otimes (\hat{v}_k - v'_k) + (\hat{v}_k - v'_k) \otimes v'_k\| \\
&\leq 2 \|P\| \lambda_K^{-1} \sum_{k=1}^K \|\hat{v}_k - v'_k\| \\
&\leq 2\lambda_K^{-1} \|P\| \left\| \hat{Q} - Q \right\| \sum_{k=1}^K \tau_k.
\end{aligned}$$

Theorem 2 and the assumption that $\ln T \left(\sum_{i=1}^K \tau_k \right)^2 / T\lambda_K^2 \rightarrow 0$ then imply that

$$\|D_3\| \rightarrow 0 \quad \text{a.s.}$$

as $T \rightarrow \infty$.

In the end, notice that $Av_k = P\lambda_k^{-1}v_k$, then

$$\begin{aligned}
\|D_4\| &= \sup_{\|x\| \leq 1} \left| P \sum_{k=1}^K \lambda_k^{-1} (v_k \otimes v_k)(x) - A(x) \right| \\
&= \sup_{\{\alpha_k: \sum_{k=1}^{\infty} \alpha_k^2 \leq 1\}} \left| P \sum_{k=1}^K \lambda_k^{-1} (v_k \otimes v_k) \left(\sum_{k=1}^{\infty} \alpha_k v_k \right) - A \left(\sum_{k=1}^{\infty} \alpha_k v_k \right) \right| \\
&= \sup_{\{\alpha_k: \sum_{k=1}^{\infty} \alpha_k^2 \leq 1\}} \left| \sum_{k=1}^K \alpha_k P \lambda_k^{-1} v_k - \sum_{k=1}^{\infty} \alpha_k Av_k \right| \\
&= \sup_{\{\alpha_k: \sum_{k=1}^{\infty} \alpha_k^2 \leq 1\}} \left| \sum_{k=1}^K \alpha_k Av_k - \sum_{k=1}^{\infty} \alpha_k Av_k \right| \\
&= \sup_{\{\alpha_k: \sum_{k=1}^{\infty} \alpha_k^2 \leq 1\}} \left| \sum_{k>K}^{\infty} \alpha_k Av_k \right| \\
&\leq \sup_{\{\alpha_k: \sum_{k=1}^{\infty} \alpha_k^2 \leq 1\}} \left(\sum_{k>K}^{\infty} \alpha_k^2 \right)^{1/2} \left(\sum_{k>K}^{\infty} |Av_k|^2 \right)^{1/2} \\
&\leq \left(\sum_{k>K}^{\infty} |Av_k|^2 \right)^{1/2}.
\end{aligned}$$

Since A is Hilbert-Schmidt, $\sum_{k=1}^{\infty} |Av_k|^2 < \infty$. This implies that $\sum_{k>K}^{\infty} |Av_k|^2 \rightarrow 0$ as $T \rightarrow \infty$. Then

$$\|D_4\| \rightarrow 0$$

as $T \rightarrow \infty$. Then

$$\left\| \hat{A}_K - A \right\| \rightarrow 0 \quad \text{a.s.}$$

as $T \rightarrow \infty$. □

Proof of Corollary 6. By Theorem 5, we have that $\left\| \hat{A}_k - A \right\| \rightarrow 0$ a.s. as $T \rightarrow \infty$. This implies that

$$\left\| \hat{A}_K \right\| = O(1) \quad \text{a.s..}$$

In the proof of Theorem 2, we showed that $\|w_t\| \leq 2M$ a.s. and that $\|\hat{w}_t\| \leq 4M$ a.s.. Consequently, we have that

$$\|\varepsilon_t\| = \|w_t - Aw_{t-1}\| \leq \|w_t\| + \|A\| \|w_{t-1}\| = O(1) \quad \text{a.s..}$$

and that

$$\|\hat{\varepsilon}_t\| = \left\| \hat{w}_t - \hat{A}_K \hat{w}_{t-1} \right\| \leq \|\hat{w}_t\| + \left\| \hat{A}_K \right\| \|\hat{w}_{t-1}\| = O(1) \quad \text{a.s..}$$

Also,

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \|\hat{\varepsilon}_t - \varepsilon_t\| &= \frac{1}{T} \sum_{t=1}^T \left\| (\hat{w}_t - w_t) - \hat{A}_K(\hat{w}_{t-1} - w_{t-1}) + (A - \hat{A}_K)w_{t-1} \right\| \\ &\leq \frac{1}{T} \sum_{t=1}^T \|\hat{w}_t - w_t\| + \frac{\left\| \hat{A}_K \right\|}{T} \sum_{t=1}^T \|\hat{w}_{t-1} - w_{t-1}\| + \frac{\left\| A - \hat{A}_K \right\|}{T} \sum_{t=1}^T \|w_{t-1}\| \\ &\rightarrow 0 \quad \text{a.s.} \end{aligned}$$

by (20) and Theorem 5. Then,

$$\begin{aligned} \left\| \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t \otimes \hat{\varepsilon}_t - \frac{1}{T} \sum_{t=1}^T \varepsilon_t \otimes \varepsilon_t \right\| &= \left\| \frac{1}{T} \sum_{t=1}^T [\hat{\varepsilon}_t \otimes (\hat{\varepsilon}_t - \varepsilon_t) + (\hat{\varepsilon}_t - \varepsilon_t) \otimes \varepsilon_t] \right\| \\ &\leq \frac{1}{T} \sum_{t=1}^T (\|\hat{\varepsilon}_t\| + \|\varepsilon_t\|) \|\hat{\varepsilon}_t - \varepsilon_t\| \\ &\rightarrow 0 \quad \text{a.s..} \end{aligned}$$

By Theorem 2.4 in Bosq (2000),

$$\left\| \frac{1}{T} \sum_{t=1}^T \varepsilon_t \otimes \varepsilon_t - \mathbb{E}(\varepsilon_t \otimes \varepsilon_t) \right\| \rightarrow 0 \quad \text{a.s.}$$

as $T \rightarrow \infty$. Then

$$\begin{aligned} \left\| \hat{\Sigma} - \Sigma \right\| &\leq \left\| \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t \otimes \hat{\varepsilon}_t - \frac{1}{T} \sum_{t=1}^T \varepsilon_t \otimes \varepsilon_t \right\| + \left\| \frac{1}{T} \sum_{t=1}^T \varepsilon_t \otimes \varepsilon_t - \mathbb{E}(\varepsilon_t \otimes \varepsilon_t) \right\| \\ &\rightarrow 0 \quad \text{a.s.} \end{aligned}$$

as $T \rightarrow \infty$. □

Proof of Lemma 7. Write

$$\hat{w}_t = A\hat{w}_{t-1} + \varepsilon_t + (\hat{w}_t - w_t) - A(\hat{w}_{t-1} - w_{t-1}),$$

from which we may deduce that

$$\hat{P} = A\hat{Q} + S + (R_1 + R_2 + R_3), \tag{25}$$

where

$$S = \frac{1}{T} \sum_{t=1}^T \varepsilon_t \otimes w_{t-1},$$

and

$$\begin{aligned} R_1 &= \frac{1}{T} \sum_{t=1}^T \varepsilon_t \otimes (\hat{w}_{t-1} - w_{t-1}), \\ R_2 &= \frac{1}{T} \sum_{t=1}^T (\hat{w}_t - w_t) \otimes \hat{w}_{t-1}, \\ R_3 &= - \left[\frac{1}{T} \sum_{t=1}^T A(\hat{w}_{t-1} - w_{t-1}) \otimes \hat{w}_{t-1} \right]. \end{aligned}$$

Since

$$\begin{aligned} \hat{A}_K &= \hat{P}\hat{Q}_K^+ \\ &= A\hat{Q}\hat{Q}_K^+ + S\hat{Q}_K^+ + (R_1 + R_2 + R_3)\hat{Q}_K^+ \\ &= A\hat{\Pi}_K + S(\hat{Q}_K^+ - Q_K^+) + SQ_K^+ + (R_1 + R_2 + R_3)\hat{Q}_K^+, \end{aligned}$$

we have that

$$\left(\hat{A}_K - A\hat{\Pi}_K\right)(w_T) = SQ_K^+(w_T) + (R_1 + R_2 + R_3)\hat{Q}_K^+(w_T) + S(\hat{Q}_K^+ - Q_K^+)(w_T).$$

It has been established in Mas (2007) that

$$\sqrt{T/K}SQ_K^+(w_T) \rightarrow_d \mathbb{N}(0, \Sigma).$$

To show that

$$\sqrt{T/K}\left(\hat{A}_K\hat{w}_T - A\hat{\Pi}_K w_T\right) \rightarrow_d \mathbb{N}(0, \Sigma), \quad (26)$$

it suffices to show that

$$\sqrt{T/K}(R_1 + R_2 + R_3)\hat{Q}_K^+(w_T) \rightarrow_p 0, \quad (27)$$

$$\sqrt{T/K}S(\hat{Q}_K^+ - Q_K^+)(w_T) \rightarrow_p 0, \quad (28)$$

and that

$$\sqrt{T/K}\hat{A}_K(\hat{w}_T - w_T) \rightarrow_p 0. \quad (29)$$

To show that (27) holds, we first note that for R_1 , we have that

$$\begin{aligned} \|R_1\| &= \left\| \frac{1}{T} \sum_{t=1}^T \varepsilon_t \otimes [\Delta_t - (\bar{f} - \mathbb{E}f)] \right\| \\ &\leq \left\| \frac{1}{T} \sum_{t=1}^T \varepsilon_t \otimes \Delta_t \right\| + \left\| \frac{1}{T} \sum_{t=1}^T \varepsilon_t \right\| \|(\bar{f} - \mathbb{E}f)\|. \end{aligned}$$

Since $\mathbb{E}\|\varepsilon_t\|^2 < \infty$, and since that $\sup_{t \geq 1} \mathbb{E}\|\Delta_t\|^2 = O(N^{-r})$ and $N \geq cT^{2/r} \ln^s T$ for some constant $c > 0$ and $s > 0$ together imply that $\sup_{t \geq 1} \mathbb{E}\|\Delta_t\|^2 = o(T^{-2})$, we have

$$\begin{aligned} \mathbb{E} \left\| \frac{1}{T} \sum_{t=1}^T \varepsilon_t \otimes \Delta_t \right\| &\leq \frac{1}{T} \sum_{t=1}^T \mathbb{E} \|\varepsilon_t\| \|\Delta_t\| \\ &\leq \frac{1}{T} \sum_{t=1}^T \left(\mathbb{E} \|\varepsilon_t\|^2\right)^{1/2} \left(\mathbb{E} \|\Delta_t\|^2\right)^{1/2} \\ &\leq \left(\mathbb{E} \|\varepsilon_t\|^2\right)^{1/2} \left(\sup_{t \geq 1} \mathbb{E} \|\Delta_t\|^2\right)^{1/2} = o(T^{-1}). \end{aligned}$$

This implies that

$$\left\| \frac{1}{T} \sum_{t=1}^T \varepsilon_t \otimes \Delta_t \right\| = o_p(T^{-1}).$$

By Theorem 2.7 in Bosq (2000) we have that

$$\left\| \frac{1}{T} \sum_{t=1}^T \varepsilon_t \right\| = O_p(T^{-1/2}).$$

Also we have proved in Lemma 1 that $\|\bar{f} - \mathbb{E}f\| = O(T^{-1/2} \ln^{1/2} T)$ a.s.. Then we have that $\|R_1\| = O_p(T^{-1} \ln^{1/2} T)$.

For R_2 , write

$$R_2 = \frac{1}{T} \sum_{t=1}^T (\hat{w}_t - w_t) \otimes w_{t-1} + \frac{1}{T} \sum_{t=1}^T (\hat{w}_t - w_t) \otimes (\hat{w}_{t-1} - w_{t-1}). \quad (30)$$

For the first term on the right hand side of (30), we have that

$$\left\| \frac{1}{T} \sum_{t=1}^T (\hat{w}_t - w_t) \otimes w_{t-1} \right\| \leq \left\| \frac{1}{T} \sum_{t=1}^T \Delta_t \otimes w_{t-1} \right\| + \|\bar{f} - \mathbb{E}f\| \left\| \frac{1}{T} \sum_{t=1}^T w_{t-1} \right\|.$$

Since $\|w_t\|^2 \leq 4M^2$ a.s. for all t , we have that

$$\begin{aligned} \mathbb{E} \left\| \frac{1}{T} \sum_{t=1}^T \Delta_t \otimes w_{t-1} \right\| &\leq \frac{1}{T} \sum_{t=1}^T \mathbb{E} \|\Delta_t\| \|w_{t-1}\| \\ &\leq \frac{1}{T} \sum_{t=1}^T \left(\mathbb{E} \|\Delta_t\|^2 \right)^{1/2} \left(\mathbb{E} \|w_{t-1}\|^2 \right)^{1/2} \\ &\leq 2M \left(\sup_{k \geq 1} \mathbb{E} \|\Delta_k\|^2 \right)^{1/2} = o(T^{-1}). \end{aligned} \quad (31)$$

This implies that

$$\left\| \frac{1}{T} \sum_{t=1}^T \Delta_t \otimes w_{t-1} \right\| = o_p(T^{-1}).$$

By Corollary 3.2 in Bosq (2000), we have that

$$\left\| \frac{1}{T} \sum_{t=1}^T w_{t-1} \right\| = O(T^{-1/2} \ln^{1/2} T) \quad \text{a.s..}$$

By Lemma 1, we have that $\|\bar{f} - \mathbb{E}f\| = O(T^{-1/2} \ln^{1/2} T)$ a.s.. Then

$$\left\| \frac{1}{T} \sum_{t=1}^T (\hat{w}_t - w_t) \otimes w_{t-1} \right\| = O_p(T^{-1} \ln T).$$

For the second term on the right hand side of (30), notice that by (19), we have

$$\begin{aligned} \mathbb{E} \left\| \frac{1}{T} \sum_{t=1}^T (\hat{w}_t - w_t) \otimes (\hat{w}_{t-1} - w_{t-1}) \right\| &\leq \frac{1}{T} \sum_{t=1}^T \left(\mathbb{E} \|\hat{w}_t - w_t\|^2 \right)^{1/2} \left(\mathbb{E} \|\hat{w}_{t-1} - w_{t-1}\|^2 \right)^{1/2} \\ &= O(T^{-1}). \end{aligned}$$

This implies that

$$\left\| \frac{1}{T} \sum_{t=1}^T (\hat{w}_t - w_t) \otimes (\hat{w}_{t-1} - w_{t-1}) \right\| = O_p(T^{-1}).$$

Then we have that $\|R_2\| = O_p(T^{-1} \ln T)$.

Similarly, one can show that $\|R_3\| = O_p(T^{-1} \ln T)$. Then $\|R_1 + R_2 + R_3\| = O_p(T^{-1} \ln T)$. Together with (21) and that $\|w_T\| \leq 2M$ a.s., this implies that

$$\sqrt{T/K} \left\| (R_1 + R_2 + R_3) \hat{Q}_K^+(w_T) \right\| \leq \sqrt{T/K} \|R_1 + R_2 + R_3\| \lambda_K^{-1} \|w_T\| \rightarrow_p 0.$$

Then (27) holds.

To prove that (28) holds, note that

$$\left\| \hat{Q}_K^+ - Q_K^+ \right\| \leq \|E_2\| + \|E_3\|$$

where

$$E_2 = \sum_{k=1}^K \hat{\lambda}_k^{-1} (\hat{v}_k \otimes \hat{v}_k) - \sum_{k=1}^K \lambda_k^{-1} (\hat{v}_k \otimes \hat{v}_k),$$

and

$$E_3 = \sum_{k=1}^K \lambda_k^{-1} (\hat{v}_k \otimes \hat{v}_k) - \sum_{k=1}^K \lambda_k^{-1} (v_k \otimes v_k).$$

That $\|E_2\| \rightarrow_p 0$ and $\|E_3\| \rightarrow_p 0$ follow immediately from the proofs of $\|D_2\| \rightarrow_{a.s.} 0$ and $\|D_3\| \rightarrow_{a.s.} 0$ in Theorem 5. Therefore, $\left\| \hat{Q}_K^+ - Q_K^+ \right\| \rightarrow_p 0$.

It follows from the proof of Lemma 4.6 in Bosq (2000) that $\|S\| = O_p(1/\sqrt{T})$. Then

$$\sqrt{T/K} \left\| S(\hat{Q}_K^+ - Q_K^+)(w_T) \right\| \leq \sqrt{T/K} \|S\| \left\| \hat{Q}_K^+ - Q_K^+ \right\| \|w_T\| \rightarrow_p 0.$$

That is, (28) holds.

By (19), $\|\hat{w}_T - w_T\| = O_p(1/\sqrt{T})$. Then (29) follows immediately. This completes the proof of the Lemma. \square

Proof of Theorem 8. Due to results in Lemma 7, it suffices to show that

$$\sqrt{T/K}(A\hat{\Pi}_K w_T - Aw_T) = o_p(1).$$

Since

$$\begin{aligned} \left\| A\hat{\Pi}_K w_T - Aw_T \right\| &\leq \left\| A\hat{\Pi}_K w_T - A\Pi_K w_T \right\| + \left\| A\Pi_K w_T - Aw_T \right\| \\ &\leq \|A\| \left\| \hat{\Pi}_K - \Pi_K \right\| \|w_T\| + \|A\| \|(I - \Pi_K)w_T\|, \end{aligned}$$

it suffices to show that

$$\left\| \hat{\Pi}_K - \Pi_K \right\| = o_p\left(T^{-1/2}K^{1/2}\right) \quad (32)$$

and that

$$\|(I - \Pi_K)w_T\| = o_p\left(T^{-1/2}K^{1/2}\right). \quad (33)$$

To prove (32), we need some tools from functional calculus. Let

$$\delta_k = \min\{\lambda_k - \lambda_{k+1}, \lambda_{k-1} - \lambda_k\}$$

and

$$\hat{\delta}_k = \min\{\hat{\lambda}_k - \hat{\lambda}_{k+1}, \hat{\lambda}_{k-1} - \hat{\lambda}_k\}$$

where we set $\lambda_0 = \hat{\lambda}_0 = +\infty$. Notice that by convexity of λ_k , we have that $\delta_k = \lambda_k - \lambda_{k+1}$. Let B_k be the oriented circle in the complex plane with center λ_k and radius $\delta_k/3$ and let \hat{B}_k be the oriented circle in the complex plane with center $\hat{\lambda}_k$ and radius $\hat{\delta}_k/3$. Notice that the B_k 's are disjoint. Also, the \hat{B}_k 's are disjoint. Let $B = \bigcup_{k=1}^K B_k$ and $\hat{B} = \bigcup_{k=1}^K \hat{B}_k$. Then we have that

$$\Pi_K = \frac{1}{2\pi i} \oint_B (zI - Q)^{-1} dz$$

and

$$\hat{\Pi}_K = \frac{1}{2\pi i} \oint_{\hat{B}} (zI - \hat{Q})^{-1} dz. \quad (34)$$

The contour integrals here are defined as Stieltjes integrals on the complex plane whose convergences are in the norm of $\mathcal{L}(H)$, the space of all bounded linear operators on H . For details of functional calculus, one may see, for example, Gohberg et al. (1990, Chapter 1).

Next, we show that

$$\frac{\sqrt{T}}{2\pi i} \oint_{\hat{B}} (zI - \hat{Q})^{-1} dz \rightarrow_p \frac{\sqrt{T}}{2\pi i} \oint_B (zI - \hat{Q})^{-1} dz. \quad (35)$$

To show (35), we define

$$\mathcal{A}_T = \bigcap_{k=1}^K \left\{ \omega \in \Omega : \left| \hat{\lambda}_k(\omega) - \lambda_k \right| < \frac{\delta_k}{3} \right\}.$$

Note that for all $\omega \in \mathcal{A}_T$, $\hat{\lambda}_k(\omega)$ lies in the interior of the circle B_k for all $k \in \{1, 2, \dots, K\}$. This implies that we may replace all \hat{B}_k 's with B_k 's and therefore replace \hat{B} with B in (34) without affecting the value of the right hand side of (34) for all $\omega \in \mathcal{A}_T$. Then

$$\begin{aligned} \frac{\sqrt{T}}{2\pi i} \oint_{\hat{B}} (zI - \hat{Q})^{-1} dz &= \frac{\sqrt{T}}{2\pi i} \oint_B (zI - \hat{Q})^{-1} dz \\ &+ \left(\frac{\sqrt{T}}{2\pi i} \oint_{\hat{B}} (zI - \hat{Q})^{-1} dz - \frac{\sqrt{T}}{2\pi i} \oint_B (zI - \hat{Q})^{-1} dz \right) \mathbf{1}(\mathcal{A}_T^C). \end{aligned}$$

Denote

$$L_1 = \left(\frac{\sqrt{T}}{2\pi i} \oint_{\hat{B}} (zI - \hat{Q})^{-1} dz - \frac{\sqrt{T}}{2\pi i} \oint_B (zI - \hat{Q})^{-1} dz \right) \mathbf{1}(\mathcal{A}_T^C).$$

By Theorem 4.10 in Bosq (2000), the asymptotic behavior of $|\hat{\lambda}_k - \lambda_k|$ in distribution is the same as that of $\langle (\hat{Q} - Q)v_k, v_k \rangle$. Then for any $\eta > 0$, we have that

$$\begin{aligned} \mathbb{P}(\|L_1\| > \eta) &\leq \mathbb{P}(\mathcal{A}_T^C) \leq \sum_{k=1}^K \mathbb{P}\left(\left| \hat{\lambda}_k - \lambda_k \right| \geq \frac{\delta_k}{3} \right) \\ &\leq \frac{3\lambda_k}{\delta_k} \mathbb{E} \frac{\left| \hat{\lambda}_k - \lambda_k \right|}{\lambda_k} \\ &\sim \frac{3\lambda_k}{\delta_k} \mathbb{E} \frac{\langle (\hat{Q} - Q)v_k, v_k \rangle}{\lambda_k}. \end{aligned}$$

To prove that the above term converges to 0, we first prove that

$$\sup_{m,p} T \frac{\mathbb{E} \langle (\hat{Q} - Q)v_p, v_m \rangle^2}{\lambda_p \lambda_m} \leq M \quad (36)$$

where M is some constant. Write

$$\hat{Q} - Q = \tilde{Q} - Q + R,$$

where \tilde{Q} has been defined in Theorem 2 and R includes all residual terms. Then

$$\langle (\hat{Q} - Q)v_p, v_m \rangle^2 = \langle (\tilde{Q} - Q)v_p, v_m \rangle^2 + \langle Rv_p, v_m \rangle^2 + 2\langle (\tilde{Q} - Q)v_p, v_m \rangle \langle Rv_p, v_m \rangle.$$

It has been established in Mas (2007) that

$$\sup_{m,p} T \frac{\mathbb{E} \langle (\tilde{Q} - Q)v_p, v_m \rangle^2}{\lambda_p \lambda_m} \leq M.$$

Other terms are negligible. Here we examine $\langle Rv_p, v_m \rangle^2$. For example, a typical term of R is $R_0 = T^{-1} \sum_{t=1}^T \Delta_t \otimes w_t$. We have

$$\begin{aligned} \langle R_0 v_p, v_m \rangle^2 &= \frac{1}{T^2} \left(\sum_{t=1}^T \langle w_t, v_p \rangle \langle \Delta_t, v_m \rangle \right)^2 \\ &= \frac{1}{T^2} \sum_{t=1}^T \langle w_t, v_p \rangle^2 \langle \Delta_t, v_m \rangle^2 + \frac{2}{T^2} \sum_{1 \leq t < s \leq T} \langle w_t, v_p \rangle \langle \Delta_t, v_m \rangle \langle w_s, v_p \rangle \langle \Delta_s, v_m \rangle. \end{aligned}$$

We have

$$\begin{aligned} \mathbb{E} \frac{1}{T^2} \sum_{t=1}^T \langle w_t, v_p \rangle^2 \langle \Delta_t, v_m \rangle^2 &\leq \frac{1}{T^2} \sum_{t=1}^T (\mathbb{E} \langle w_t, v_p \rangle^4)^{1/2} (\mathbb{E} \langle \Delta_t, v_m \rangle^4)^{1/2} \\ &\leq \frac{M \lambda_m \lambda_p}{T} \sup_m \frac{\mathbb{E} \langle \Delta_1, v_m \rangle^2}{\lambda_m} = \frac{M \lambda_p \lambda_m}{T} O(N^{-r}). \end{aligned}$$

Also, note that

$$\begin{aligned}
& \left| \mathbb{E} \frac{2}{T^2} \sum_{1 \leq t < s \leq T} \langle w_t, v_p \rangle \langle \Delta_t, v_m \rangle \langle w_s, v_p \rangle \langle \Delta_s, v_m \rangle \right| \\
& \leq \frac{2}{T^2} \sum_{1 \leq t < s \leq T} M b^{|t-s|} \left[\mathbb{E} (\langle w_t, v_p \rangle^2 \langle w_s, v_p \rangle^2) \mathbb{E} (\langle \Delta_t, v_m \rangle^2 \langle \Delta_s, v_m \rangle^2) \right]^{1/2} \\
& \leq \frac{2}{T^2} \sum_{1 \leq t < s \leq T} M b^{|t-s|} \left[(\mathbb{E} \langle w_t, v_p \rangle^4) (\mathbb{E} \langle w_s, v_p \rangle^4) (\mathbb{E} \langle \Delta_t, v_m \rangle^4) (\mathbb{E} \langle \Delta_s, v_m \rangle^4) \right]^{1/2} \\
& \leq \frac{\lambda_m \lambda_p}{T} \sum_{j=1}^{T-1} \left(1 - \frac{j}{T} \right) b^j O(N^{-r}) = \frac{\lambda_m \lambda_p}{T} O(N^{-r}),
\end{aligned}$$

Similar arguments apply to other terms in R . Now we may conclude (36) and therefore that

$$\frac{3\lambda_k}{\delta_k} \mathbb{E} \frac{\langle (\hat{Q} - Q)v_k, v_k \rangle}{\lambda_k} \leq \frac{M}{\sqrt{T}} \sum_{k=1}^K \frac{3\lambda_k}{\delta_k} \leq \frac{3M}{\sqrt{T}} \sum_{k=1}^K k \ln k \leq \frac{3M}{\sqrt{T}} K^2 \ln K.$$

By assumption, the last term converges to 0. This implies that $\lim_{T \rightarrow \infty} \mathbb{P}(\|L_1\| > \eta) = 0$. Then (35) holds. It then implies that

$$\begin{aligned}
\sqrt{T/K} (\hat{\Pi}_K - \Pi_K) & \rightarrow_p \sqrt{\frac{T}{K}} \left(\frac{1}{2\pi i} \oint_B [(zI - \hat{Q})^{-1} - (zI - Q)^{-1}] dz \right) \\
& = \sqrt{\frac{T}{K}} \left(\frac{1}{2\pi i} \oint_B [(zI - Q)^{-1} (\hat{Q} - Q) (zI - \hat{Q})^{-1}] dz \right).
\end{aligned}$$

Now consider

$$L_2 = (zI - Q)^{-1} (\hat{Q} - Q) (zI - \hat{Q})^{-1} - (zI - Q)^{-1} (\hat{Q} - Q) (zI - Q)^{-1}$$

and write

$$\begin{aligned}
L_2 & = (zI - Q)^{-1} (\hat{Q} - Q) [(zI - \hat{Q})^{-1} - (zI - Q)^{-1}] \\
& = (zI - Q)^{-1} (\hat{Q} - Q) (zI - Q)^{-1} (\hat{Q} - Q) (zI - \hat{Q})^{-1} \\
& = (zI - Q)^{-\frac{1}{2}} \left[(zI - Q)^{-\frac{1}{2}} (\hat{Q} - Q) (zI - Q)^{-\frac{1}{2}} \right]^2 \\
& \quad \cdot \left[(zI - Q)^{\frac{1}{2}} (zI - \hat{Q}) (zI - Q)^{\frac{1}{2}} \right] (zI - Q)^{-\frac{1}{2}}.
\end{aligned} \tag{37}$$

With minor modifications of proofs in Mas (2007), we may show that for

$$\mathcal{E}_k = \left\{ \sup_{z \in B_k} \left\| (zI - Q)^{-\frac{1}{2}} (\hat{Q} - Q) (zI - Q)^{-\frac{1}{2}} \right\| < \frac{1}{2} \right\},$$

we have that

$$\sup_{z \in B_k} \left\| (zI - Q)^{\frac{1}{2}} (zI - \hat{Q}) (zI - Q)^{\frac{1}{2}} \right\| 1(\mathcal{E}_k) \leq 2 \quad \text{a.s.} \quad (38)$$

and that

$$\mathbb{P}(\mathcal{E}_k^C) \leq M \frac{k \ln k}{\sqrt{T}}. \quad (39)$$

Also, for k large enough, we have that

$$\mathbb{E} \sup_{z \in B_k} \left\| (zI - Q)^{-\frac{1}{2}} (\hat{Q} - Q) (zI - Q)^{-\frac{1}{2}} \right\|^2 \leq \frac{M}{T} k^2 \ln^2 k. \quad (40)$$

Equation (39) implies that for any $\eta > 0$,

$$\mathbb{P} \left(\sqrt{\frac{T}{K}} \frac{1}{2\pi i} \oint_B L_2 1 \left(\bigcup_{k=1}^K \mathcal{E}_k^C \right) dz > \eta \right) \leq \sum_{k=1}^K \mathbb{P}(\mathcal{E}_k^C) \leq \frac{M}{\sqrt{T}} K^2 \ln K \rightarrow 0 \quad (41)$$

under the assumption. On the other hand, (37), (38), (40), and the fact that $\|zI - Q\|^{-1/2} \leq \delta_k^{-1/2}$ for all $k \in \{1, 2, \dots, K\}$ implies that

$$\begin{aligned} \sqrt{\frac{T}{K}} \frac{1}{2\pi i} \oint_B L_2 1 \left(\bigcap_{k=1}^K \mathcal{E}_k \right) dz &= \sqrt{\frac{T}{K}} \frac{1}{2\pi i} \sum_{k=1}^K \oint_{B_k} L_2 1(\mathcal{E}_k) dz \\ &\leq \sqrt{\frac{T}{K}} \frac{1}{2\pi i} \sum_{k=1}^K \oint_{B_k} \delta_k^{-1/2} \frac{M}{T} k^2 \ln^2 k \cdot 2\delta_k^{-1/2} dz \quad \text{a.s.} \\ &\leq \frac{M'}{\sqrt{T}} K^{5/2} \ln^2 K \quad \text{a.s..} \end{aligned}$$

By assumption, the last term converges to 0. This, together with (41) implies that

$$\sqrt{\frac{T}{K}} \frac{1}{2\pi i} \oint_B L_2 dz \rightarrow_p 0,$$

that is,

$$\sqrt{T/K} \left(\hat{\Pi}_K - \Pi_K \right) \rightarrow_p \sqrt{\frac{T}{K}} \left(\frac{1}{2\pi i} \oint_B \left[(zI - Q)^{-1} (\hat{Q} - Q) (zI - Q)^{-1} \right] dz \right).$$

It is shown in Dauxois et al. (1982) that

$$\frac{1}{2\pi i} \oint_{B_k} (zI - Q)^{-1} (\hat{Q} - Q) (zI - Q)^{-1} dz = \psi_k (\hat{Q} - Q) \pi_k + \pi_k (\hat{Q} - Q) \psi_k$$

where $\pi_k = v_k \otimes v_k$ is the projection on the eigenspace associated with λ_k , and $\psi_k = \sum_{\ell \neq k} \pi_\ell / (\lambda_\ell - \lambda_k)$. As a consequence, we have that

$$\langle v_i, (\hat{\Pi}_K - \Pi_K)v_j \rangle = \begin{cases} \frac{\langle v_i, (\hat{Q} - Q)v_j \rangle}{\lambda_i - \lambda_j}, & 1 \leq i \leq K < j, \\ \frac{\langle v_i, (\hat{Q} - Q)v_j \rangle}{\lambda_j - \lambda_i}, & 1 \leq j \leq K < i, \end{cases}$$

with all other terms vanishing.

Then

$$\begin{aligned} \left\| \hat{\Pi}_K - \Pi_K \right\| &= \sup_{\|v\|=1} \langle v, (\hat{\Pi}_K - \Pi_K)v \rangle \\ &= \sup_{\|v\|=1} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \langle v, v_i \rangle \langle v, v_j \rangle \langle v_i, (\hat{\Pi}_K - \Pi_K)v_j \rangle \\ &= 2 \sup_{\|v\|=1} \sum_{i=1}^K \sum_{j=K+1}^{\infty} \langle v, v_i \rangle \langle v, v_j \rangle \frac{\langle v_i, (\hat{Q} - Q)v_j \rangle}{\lambda_i - \lambda_j} \\ &= 2 \sup_{\|v\|=1} \sum_{i=1}^K \sum_{j=K+1}^{\infty} \langle v, v_i \rangle \langle v, v_j \rangle \frac{\langle v_i, (\tilde{Q} - Q)v_j \rangle + \langle v_i, Rv_j \rangle}{\lambda_i - \lambda_j}. \end{aligned}$$

We write

$$\begin{aligned} &\sup_{\|v\|=1} \sum_{i=1}^K \sum_{j=K+1}^{\infty} \langle v, v_i \rangle \langle v, v_j \rangle \frac{\langle v_i, (\tilde{Q} - Q)v_j \rangle}{\lambda_i - \lambda_j} \\ &= \sup_{\|v\|=1} \sum_{i=1}^K \sum_{j=K+1}^{\infty} \langle v, v_i \rangle \langle v, v_j \rangle \frac{1}{\lambda_i - \lambda_j} \frac{1}{T} \sum_{t=1}^T \langle v_i, (w_t \otimes w_t)v_j \rangle \\ &= \sup_{\|v\|=1} \sum_{i=1}^K \sum_{j=K+1}^{\infty} \langle v, v_i \rangle \langle v, v_j \rangle \frac{\sqrt{\lambda_i \lambda_j}}{\lambda_i - \lambda_j} \frac{1}{T} \sum_{t=1}^T \frac{\langle v_i, w_t \rangle}{\sqrt{\lambda_i}} \frac{\langle v_j, w_t \rangle}{\sqrt{\lambda_j}} \\ &= \sup_{\|v\|=1} \sum_{i=1}^K \sum_{j=K+1}^{\infty} \langle v, v_i \rangle \langle v, v_j \rangle \frac{\sqrt{\lambda_i \lambda_j}}{\lambda_i - \lambda_j} \frac{1}{T} \sum_{t=1}^T m_{ijt}, \end{aligned}$$

where we have set

$$m_{ijt} = m_{it} m_{jt} = \frac{\langle v_i, w_t \rangle}{\sqrt{\lambda_i}} \frac{\langle v_j, w_t \rangle}{\sqrt{\lambda_j}}.$$

For all i and j such that $i \neq j$, $(m_{ijt})_{t \geq 1}$ is a mean zero stationary process. Moreover,

$$\mathbb{E}m_{ijt}^2 \leq (\mathbb{E}m_{it}^4)^{1/2} (\mathbb{E}m_{jt}^4)^{1/2} \leq \left(\mathbb{E} \frac{\langle v_i, w_t \rangle^4}{\lambda_i^2} \right)^{1/2} \left(\mathbb{E} \frac{\langle v_j, w_t \rangle^4}{\lambda_j^2} \right)^{1/2} \leq M,$$

and the autocovariance function Γ_{ij} of (m_{ijt}) is absolutely summable. As a result, we have that

$$T \mathbb{E} \left(\frac{1}{T} \sum_{t=1}^T m_{ijt} \right)^2 = \frac{1}{T} \mathbb{E} \sum_{t=1}^T \sum_{s=1}^T m_{ijt} m_{ijs} = \sum_{k=1}^T \left(1 - \frac{k}{T} \right) \Gamma_{ij}(k) < \infty$$

for all $i \neq j$. Therefore, we have that

$$\left(\frac{1}{T} \sum_{t=1}^T m_{ijt} \right)^2 = O_p(T^{-1})$$

uniformly for all $i = 1, \dots, K$ and $j = K+1, K+2, \dots$. Obviously, for any $\|v\| = 1$, we have that

$$\sum_{i=1}^K \sum_{j=K+1}^{\infty} \langle v, v_i \rangle^2 \langle v, v_j \rangle^2 \leq \sum_{i=1}^K \langle v, v_i \rangle^2 \sum_{j=K+1}^{\infty} \langle v, v_j \rangle^2 \leq 1.$$

Therefore, we have that

$$\begin{aligned} & \left| \sup_{\|v\|=1} \sum_{i=1}^K \sum_{j=K+1}^{\infty} \langle v, v_i \rangle \langle v, v_j \rangle \frac{\sqrt{\lambda_i \lambda_j}}{\lambda_i - \lambda_j} \frac{1}{T} \sum_{t=1}^T m_{ijt} \right| \\ & \leq \left(\sum_{i=1}^K \sum_{j=K+1}^{\infty} \langle v, v_i \rangle^2 \langle v, v_j \rangle^2 \right)^{1/2} \left(\sum_{i=1}^K \sum_{j=K+1}^{\infty} \frac{\lambda_i \lambda_j}{(\lambda_i - \lambda_j)^2} \left(\frac{1}{T} \sum_{t=1}^T m_{ijt} \right)^2 \right)^{1/2} \quad (42) \\ & = o_p(T^{-1/2} K^{1/2}) \end{aligned}$$

by assumption. That is,

$$\sup_{\|v\|=1} \sum_{i=1}^K \sum_{j=K+1}^{\infty} \langle v, v_i \rangle \langle v, v_j \rangle \frac{\langle v_i, (\tilde{Q} - Q)v_j \rangle}{\lambda_i - \lambda_j} = o_p(T^{-1/2} K^{1/2}).$$

The term

$$\sup_{\|v\|=1} \sum_{i=1}^K \sum_{j=K+1}^{\infty} \langle v, v_i \rangle \langle v, v_j \rangle \frac{\langle v_i, Rv_j \rangle}{\lambda_i - \lambda_j}$$

can be analyzed analogously. For instance, one component of this term can be written as

$$\begin{aligned} & \sup_{\|v\|=1} \sum_{i=1}^K \sum_{j=K+1}^{\infty} \langle v, v_i \rangle \langle v, v_j \rangle \frac{1}{\lambda_i - \lambda_j} \frac{1}{T} \sum_{t=1}^T \langle v_i, (w_t \otimes \Delta_t) v_j \rangle \\ &= \sup_{\|v\|=1} \sum_{i=1}^K \sum_{j=K+1}^{\infty} \langle v, v_i \rangle \langle v, v_j \rangle \frac{\sqrt{\lambda_i \lambda_j}}{\lambda_i - \lambda_j} \frac{1}{T} \sum_{t=1}^T \frac{\langle v_i, w_t \rangle}{\sqrt{\lambda_i}} \frac{\langle v_j, \Delta_t \rangle}{\sqrt{\lambda_j}}. \end{aligned}$$

Since

$$\begin{aligned} \mathbb{E} \left| \frac{1}{T} \sum_{t=1}^T \frac{\langle v_i, w_t \rangle}{\sqrt{\lambda_i}} \frac{\langle v_j, \Delta_t \rangle}{\sqrt{\lambda_j}} \right| &\leq \mathbb{E} \left(\frac{1}{T} \sum_{t=1}^T \frac{\langle v_i, w_t \rangle^2}{\lambda_i} \right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^T \frac{\langle v_j, \Delta_t \rangle^2}{\lambda_j} \right)^{1/2} \\ &\leq \left(\frac{1}{T} \sum_{t=1}^T \frac{\mathbb{E} \langle v_i, w_t \rangle^2}{\lambda_i} \right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^T \frac{\mathbb{E} \langle v_j, \Delta_t \rangle^2}{\lambda_j} \right)^{1/2} \\ &= O(N^{-r/2}), \end{aligned}$$

a similar argument as in (42) shows that

$$\sup_{\|v\|=1} \sum_{i=1}^K \sum_{j=K+1}^{\infty} \langle v, v_i \rangle \langle v, v_j \rangle \frac{1}{\lambda_i - \lambda_j} \frac{1}{T} \sum_{t=1}^T \langle v_i, (w_t \otimes \Delta_t) v_j \rangle = o_p(T^{-1/2} K^{1/2}).$$

Similarly, the other components could be shown to have order $o_p(T^{-1/2} K^{1/2})$. Therefore, we conclude that

$$\sup_{\|v\|=1} \sum_{i=1}^K \sum_{j=K+1}^{\infty} \langle v, v_i \rangle \langle v, v_j \rangle \frac{\langle v_i, R v_j \rangle}{\lambda_i - \lambda_j} = o_p(T^{-1/2} K^{1/2}),$$

and hence

$$\sqrt{T/K} \left\| \hat{\Pi}_K - \Pi_K \right\| = o_p(1).$$

To complete the proof of the theorem, we only need to note that

$$\begin{aligned}
\mathbb{E} \|(I - \Pi_K)w_T\|^2 &= \mathbb{E} \left\| \left(\sum_{k=K+1}^{\infty} v_k \otimes v_k \right) w_T \right\|^2 \\
&= \mathbb{E} \left\| \sum_{k=K+1}^{\infty} \langle v_k, w_T \rangle v_k \right\|^2 \\
&= \mathbb{E} \left(\sum_{k=K+1}^{\infty} \langle v_k, w_T \rangle^2 \right) \\
&= \sum_{k=K+1}^{\infty} \lambda_k = o(T^{-1}K).
\end{aligned}$$

□

References

- Antoniadis, A., Paparoditis, E., Sapatinas, T., 2006. A functional wavelet–kernel approach for time series prediction. *Journal of the Royal Statistical Society. Series B (Methodological)* 68 (5), 837–857.
- Aue, A., Horváth, L., Pellatt, D. F., 2017. Functional generalized autoregressive conditional heteroskedasticity. *Journal of Time Series Analysis* 38 (1), 3–21.
- Aue, A., Norinho, D. D., Hörmann, S., 2015. On the prediction of stationary functional time series. *Journal of the American Statistical Association* 110, 378–392.
- Benatia, D., Carrasco, M., Florens, J.-P., 2017. Functional linear regression with functional response. *Journal of Econometrics* 201, 269–291.
- Besse, P. C., Cardot, H., Stephenson, D. B., 2000. Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics* 27 (4), 673–687.
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Bosq, D., 1998. *Nonparametric Statistics for Stochastic Processes*, 2nd Edition. Lecture Notes in Statistics 110. Springer-Verlag, New York.
- Bosq, D., 2000. *Linear Processes in Function Spaces: Theory and Applications*. Lecture Notes in Statistics 149. Springer-Verlag, New York.
- Brooks, C., Burke, S. P., Heravi, S., Persaud, G., 2005. Autoregressive conditional kurtosis. *Journal of Financial Econometrics* 3, 399–421.
- Dauxois, J., Pousse, A., Romain, Y., 1982. Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis* 12, 136–154.
- Didericksen, D., Kokoszka, P., Zhang, X., 2012. Empirical properties of forecasts with the functional autoregressive model. *Computational Statistics* 27 (2), 285–298.
- Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.
- Engle, R. F., Lilien, D. M., Robins, R. P., 1987. Estimating time varying risk premia in the term structure: The ARCH-M model. *Econometrica* 55, 391–407.
- Fama, E. F., French, K. R., 2012. Size, value, and momentum in international stock returns. *Journal of Financial Economics* 105 (3), 457–472.
- Ferraty, F., Vieu, P., 2006. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics. Springer, New York.

- Gohberg, I., Goldberg, S., Kaashoek, M. A., 1990. *Classes of Linear Operators*. Vol. I. Birkhäuser, Basel.
- Hansen, B. E., 1994. Autoregressive conditional density estimation. *International Economic Review* 35, 705–730.
- Harvey, C. R., Siddique, A., 1999. Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis* 34, 465–487.
- Hörmann, S., Kokoszka, P., 2010. Weakly dependent functional data. *Annals of Statistics* 38 (3), 1845–1884.
- Horváth, L., Hušková, M., Kokoszka, P., 2010. Testing the stability of the functional autoregressive process. *Journal of Multivariate Analysis* 101 (2), 352–367.
- Kargin, V., Onatski, A., 2008. Curve forecasting by functional autoregression. *Journal of Multivariate Analysis* 99, 2508–2526.
- Klepsch, J., Klüppelberg, C., 2017. An innovations algorithm for the prediction of functional linear processes. *Journal of Multivariate Analysis* 155, 252–271.
- Koenker, R., Xiao, Z., 2006. Quantile autoregression. *Journal of the American Statistical Association* 101, 980–990.
- Kokoszka, P., Reimherr, M., 2013. Determining the order of the functional autoregressive model. *Journal of Time Series Analysis* 34, 116–129.
- Mas, A., 2007. Weak convergence in the functional autoregressive model. *Journal of Multivariate Analysis* 98, 1231–1261.
- Panaretos, V. M., Tavakoli, S., 2013a. Cramér–karhunen–loève representation and harmonic principal component analysis of functional time series. *Stochastic Processes and Their Applications* 123 (7), 2779–2807.
- Panaretos, V. M., Tavakoli, S., 2013b. Fourier analysis of stationary time series in function space. *Annals of Statistics* 41 (2), 568–603.
- Ramsey, J. O., Silverman, B. W., 2005. *Functional Data Analysis*, 2nd Edition. Springer Series in Statistics. Springer-Verlag, New York.
- van den Boogaart, K. G., Egozcue, J. J., Pawlowsky-Glahn, V., 2014. Bayes Hilbert spaces. *Australian & New Zealand Journal of Statistics* 56 (2), 171–194.
- Wand, M. P., Jones, M. C., 1995. *Kernel Smoothing*. Chapman and Hall, London.
- Xiao, Z., Koenker, R., 2009. Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association* 104, 1696–1712.