Econometric Analysis of Persistent Functional Dynamics*

Yoosoon Chang†
Department of Economics
Indiana University

Bo Hu‡
Institute of New Structural Economics
Peking University

Joon Y. Park§
Department of Economics
Indiana University

This Version: June, 2023

Abstract

We introduce an autoregressive model for functional time series with unit roots. The autoregressive operator can be consistently estimated, but its convergence rate and limit distribution are different in different subspaces. In the unit root subspace, the convergence rate is fast and given by n, while the limit distribution is nonstandard and represented as functions of Brownian motions. Outside the unit root subspace, however, the limit distribution is Gaussian, although the convergence rate varies and is given by $\sqrt{n}$ or a slower rate.The predictor based on the estimated autoregressive operator has a normal limit distribution with a reduced rate of convergence. We also provide the Beveridge-Nelson decomposition, which identifies the permanent and transitory components of functional time series with unit roots, representing persistent stochastic trends and stationary cyclical movements, respectively. Using our methodology and theory, we analyze the time series of yield curves and study the dynamics of the term structure of interest rates.

JEL classification codes: C13, C14, C22, C53

Key words and phrases: functional autoregression, nonstationarity, unit root, functional Beverage-Nelson decomposition, US yield curve

*This is a substantial revision of earlier versions of this paper that have been circulated since 2016. We are grateful for comments and suggestions to Tom Sargent, Frank Diebold, Tony Smith, Richard Blundell, Luigi Pistaferri, Juan Carlos Escanciano, Doyeon Pyun, the participants of 2018 Midwest Econometrics Group Meeting, 2018 Toulouse Financial Econometrics Conference, 2018 Conference on Econometric Theory and Applications, and the seminar attendants at BI Norwegian Business School, Vanderbilt University, Boston College, Yale University, Instituto Tecnológico Autónomo de México, University of Montreal, Peking University and Indiana University. Bo Hu would like to thank NSF China for financial support (Project ID: 72003004).

†Email: yoosoon@indiana.edu
‡Email: bohu@nsd.pku.edu.cn
§Email: joon@indiana.edu

1 Introduction

The rapid development of economic theories and practices calls for econometric methods that can be used to analyze complicated objects such as curves and functions in addition to scalars and vectors. At the same time, developments in data generation, collection, storage and communication technologies give researchers access to data that have rich structures. These developments lay the groundwork for the emergence of functional data analysis in recent years, both in cross-sectional and time series settings. In functional data analysis, data are studied in the original functional form, while in traditional methods any functional observation has to be converted to a few statistics intended to summarize the information. For example, in studying distributional dynamics, one may use functional methods to keep track of the density function process, while the traditional treatments only look at the processes of a few moments and/or quantiles. Since the density function contains all the information about a distribution, functional methods provide opportunities to study full dynamics of the underlying time varying distributions, in addition to traditional methods that focus only on some particular aspects of the distributions. Functional methods therefore have advantages in studying complicated objects such as global temperature (Chang et al., 2020), electricity prices (Chen and Li, 2016), bond yield curves (Hays et al., 2012), distribution of financial returns (Hu et al., 2016; Park and Qian, 2012) and earning distribution dynamics (Chang et al., 2016a).

There is a collection of theories available for functional data analysis. Among many excellent others, Ramsey and Silverman (2005) give an introduction to the theories and tools in functional data analysis. Horväth and Kokoszka (2012) provide a comprehensive summary of the techniques in functional data analysis up to the time of publication. Ferraty and Vieu (2006) introduce nonparametric methods in functional data analysis. Bosq (2000) is devoted to the theory of functional time series, particularly functional autoregression in a stationary setting. All of these theories are developed under the assumption of independent and identical distributions or stationarity. However, many interesting functional time series in real-life applications have nonstationary features. For example, over the past 30 years, US income inequality has been growing markedly. This implies that there is likely to be nonstationarity in the density process of the US income distributions. It then calls for a framework that is able to accommodate functional time series with strong persistence. Chang et al. (2016b) give some results on functional time series with unit roots and provide a test for the number of unit roots in a functional time series. However, no formal theory has been developed for functional time series with unit roots under the autoregression setting.

In this paper, we study functional autoregression (FAR) with unit roots in infinite dimensional Hilbert spaces. We provide a functional Beveridge-Nelson decomposition that identifies the permanent and transitory components of the functional time series generated by the FAR model. These two components represent the persistent stochastic trends and stationary cyclical movements of the functional time series, respectively. We relate our decomposition to the error correction model when the underlying function space is finite dimensional. The attractor space and the cointegrating space are given by our permanent subspace and stationary subspace, respectively. We propose estimators for the functional autoregressive operator, both without and with the unit root restriction. Our estimators are consistent under very mild regularity conditions, and converge at different rates on different subspaces. In the nonstationary subspace, our estimators converge at rate n, and the limit distribution is nonstandard, given as a function of Brownian motions. Elsewhere, our estimators converge at the parametric $\sqrt{n}$-rate or at a rate slower than $\sqrt{n}$, depending on the subspaces in which the convergence is considered, and the limit distributions are Gaussian. We also provide consistent estimators for the permanent-transitory decomposition. In addition, our framework can be used to make forecasts. The one-step predictor based on our FAR estimator is asymptotically normal with a convergence rate slower than $\sqrt{n}$. We also extend our framework to incorporate the situation in which the transitory component has a non-zero drift, the data are estimated with error, and/or the functional time series is Markovian of higher order. We give conditions under which the asymptotic theory continues to hold in these extensions.

We apply our method to study the dynamics of the term structure of the US government bond yields. We model the time series of the forward rate curves by a functional autoregressive model and find that there are two dimensional unit roots in the dynamics. We decompose the forward rate curve process into its permanent and transitory components. We identify two permanent structural shocks, namely the permanent spread shock and the permanent level shock, and one transitory shock in the forward rate curve dynamics. The three shocks have at-impact effects to the forward rate curve in the forms of level change, slope change, and curvature change. We relate these three structural shocks to monetary and fiscal policy shocks, and find that the permanent spread shock and the transitory shock are related to monetary policy shocks, and the permanent level shock and the transitory shock are related to fiscal policy shocks. We get the impulse response surfaces of the yield curve to monetary and policy shocks. We find that the overall long term effect of the monetary policy shocks is significant at very short maturities, while the overall long term effect of fiscal policy shocks is significant at all maturities.

The rest of the paper is organized as follows. In Section 2 we introduce the model and the functional Beveridge-Nelson decomposition. In Section 3 we show how we may estimate the model and make prediction with the model, and develop asymptotic theories for our estimator and predictor. In Section 4 we extend our baseline model to include the case in which the stationary component has a non-zero drift, the functional time series is estimated, and/or the process is autoregressive of higher order. In Section 5 we apply our method to study the term structure of the US government bond yields. In Section 6 we present the simulation results. Section 7 concludes.

A word on exposition and notation. Our methodology and asymptotics rely heavily on a basic theory of Hilbert space, which is cited frequently throughout the paper without any specific reference. All standard notations for various notions and operations in Hilbert space are also used in the paper without any explicit definitions. The inner product and norm in our Hilbert space H are denoted as $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively, and the tensor product is defined by "$\otimes$". The superscript "$*$" is used for the adjoint of an operator on H or the dual spaces of H and its subspaces. The identity and null operator are written simply as "1" and "0". For the presentation of our estimators and their asymptotics, it would be very convenient to introduce a pseudo-inverse of a linear operator defined effectively on a proper subspace of H. For a linear transformation T defined on a proper subspace V of H, we define a pseudo-inverse T^+ of T, whenever it is well defined, to be the linear transformation such that T^+ is the inverse of T on V, and $T^+ v = 0$ for all $v \in V^\perp$, where $V^\perp$ denotes the orthogonal complement of V in H. If there is no possibility of confusion, we will simply call T^+ the inverse of T on V, or even more briefly, the inverse on V.

2 Model and Background

2.1 The Model

In the paper, we let (f_t) be a functional time series, which is regarded as a sequence of random functions taking values in a separable Hilbert space H. Formally, we may interpret f_t as an H-valued random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, i.e., $f_t : \Omega \to H$, for each $t = 1, 2, \ldots$. Throughout, we let H be given by $L^2(\mathbb{R})$, which is the Hilbert space of all square integrable real-valued functions on $\mathbb{R}$, and define $\langle u, v \rangle = \int u(r)v(r)dr$ and $\|v\| = \sqrt{\langle v, v \rangle} = (\int |v|^2(r)dr)^{1/2}$ to be the inner product and the norm defined in H, respectively. The Hilbert space $L^2(\mathbb{R})$ of square integrable functions on $\mathbb{R}$ has been used to deal with functional data in economic and financial applications. For example, Kneip and Utikal (2001) model the density functions in $L^2(\mathbb{R})$ and Kargin and Onatski (2008) analyze the Eurodollar futures rate curves in $L^2(\mathbb{R})$ with common support $[0, 1]$. Hu et al. (2016) study the dynamics of the demeaned density functions, which belongs to a subspace of $L^2(\mathbb{R})$ consisting of all functions integrated to zero with common support given by a compact subset of $\mathbb{R}$.

We suppose that the dynamics of the functional time series is given by the first-order functional autoregressive model (FAR). To be specific, we let (f_t) be generated as

$$f_t = Af_{t-1} + \varepsilon_t, \tag{1}$$

where A is a bounded linear operator on H and (ε_t) is a functional white noise whose precise meaning will be defined later. The operator norm is also denoted by $\|\cdot\|$, and therefore, we have $\|A\| = \sup_{v \in H} \|Av\| / \|v\|$. Since A is bounded, there exists a constant K such that $\|Av\| \leqslant K \|v\|$ for all $v \in H$.

The Hilbert space H is separable and admits a countable orthonormal basis. Therefore, H-valued random variables may be viewed as the infinite dimensional generalizations of random vectors. Just as an operator on a finite dimensional vector space has a matrix representation, the autoregressive operator A may be thought of as an infinite dimensional matrix with respect to any given orthonormal basis of H. In this way, the FAR may be conceptually regarded as an infinite dimensional generalization of the vector autoregression (VAR), which has been extensively used in time series econometrics. Indeed, FAR and VAR share many features. For example, just as any VAR(p) has a VAR(1) representation, any FAR(p) can be written in the FAR(1) form. This implies that the first-order Markovian assumption employed in (1) is not restrictive in any essential way. However, the introduction of infinite dimensionality does create technical difficulties. As we shall see, one problem is the lack of functional error correction representations for a very important class of functional time series with unit roots. Another issue is the so-called ill-posed inverse problem.

We first introduce some basic notions related to H-valued random variables. For an H-valued random variable f with $\mathbb{E}\|f\| < \infty$, we define its mean $\mathbb{E}f$ by the element in H such that for any $v \in H$ we have $\langle v, \mathbb{E}f \rangle = \mathbb{E}\langle v, f \rangle$.[1] Moreover, for any mean-zero H-valued random variables f and g such that $\mathbb{E}\|f\|^2 < \infty$ and $\mathbb{E}\|g\|^2 < \infty$, we define their covariance operator $\mathbb{E}(f \otimes g)$ by the operator on H such that for any u and v in H, we have $\langle u, \mathbb{E}(f \otimes g)v \rangle = \mathbb{E}\langle u, f \rangle \langle v, g \rangle$. Naturally, we call $\mathbb{E}(f \otimes f)$ the variance operator of f for any mean-zero H-valued random variable f such that $\mathbb{E}\|f\|^2 < \infty$. Using an orthonormal basis (v_k) of H, we may also define $\mathbb{E}f$ and $\mathbb{E}(f \otimes g)$ more explicitly as

$$\mathbb{E}f = \sum_{k=1}^{\infty} \big(\mathbb{E}\langle v_k, f \rangle\big) v_k \quad \text{and} \quad \mathbb{E}(f \otimes g) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \big(\mathbb{E}\langle v_i, f \rangle \langle v_j, g \rangle\big)(v_i \otimes v_j).$$

[1] Note that $\mathbb{E}\langle \cdot, f \rangle$ is a bounded linear functional on H, and therefore, $\mathbb{E}f$ exists by the Riesz representation theorem.

An H-valued white noise (ε_t) is a functional time series such that $\mathbb{E}\varepsilon_t = 0$ for all t, $\mathbb{E}(\varepsilon_t \otimes \varepsilon_t) = \Sigma$ for all t and $\mathbb{E}(\varepsilon_t \otimes \varepsilon_s) = 0$ for all $t \neq s$.

If $\|A^r\| < 1$ for some integer $r \geqslant 1$, the stochastic difference equation (1) has a stationary solution. This is shown in Bosq (2000). Stationary functional autoregressive processes have been studied by Bosq (2000) in the general setting, and by Hu et al. (2016) in a more specific setting of distributional processes with demeaned densities estimated from cross-sectional or intra-period observations.

In this paper, we consider the functional autoregressive model (1) in the presence of unit roots. Such a model is necessary to analyze the functional time series with strong persistence. Many functional time series we deal with in economic and financial applications appear to have unit roots. For instance, Chang et al. (2016b) find unit roots in the process of the density functions for the monthly cross-sectional earnings distributions in the United States, and for the intra-month S&P 500 high-frequency return distributions.

Subsequently, we denote by $\lambda(A)$ the spectrum of A, i.e., the set of all complex numbers λ such that $\lambda - A$ is not invertible on H. Note that, if H is finite dimensional, $\lambda(A)$ is the set of all eigenvalues of A. However, when H is infinite dimensional, $\lambda(A)$ is in general larger than the set of all eigenvalues of A. We assume the following throughout the paper.

Assumption 2.1. We assume that

(a) A is compact,

(b) $1 \in \lambda(A)$, and

(c) (ε_t) is independent and identically distributed with mean zero and covariance operator Σ, is independent of f_0, and $\mathbb{E}\,\|\varepsilon_t\|^4 < \infty$.

A compact operator A on H is an operator that maps the closed unit ball in H to a set whose closure is compact. It is well known that any linear operator on H is compact if and only if it can be approximated (in operator norm) by a sequence of finite rank linear operators. Part (a) of the above assumption is therefore required for a general infinite dimensional operator A to be consistently estimable by finite rank linear estimators.[2] In addition, it admits a singular value decomposition of A, which provides interesting interpretations of the dynamics in the functional process as in Hu et al. (2016). Part (b) introduces unit roots in the process (f_t). Part (c) is quite standard. The assumption of (ε_t) being independent and identically distributed with $\mathbb{E}\,\|\varepsilon_t\|^4 < \infty$ is made for simplicity, and we may readily

[2] One may potentially allow for consistently estimable non-compact operators. For example, one may set A to be determined by a finite dimensional parameter. Or one may use a sequence of non-linear operators to approximate A. However, the former approach greatly restricts the space that A lies in and the latter approach introduces non-linearity and therefore technical difficulties in inference. In view of these drawbacks, we shall stick with the compactness assumption for the autoregressive operator A.

allow (ε_t) to be a general martingale difference sequence with $\sup_{t\geqslant 1}\mathbb{E}\big(\|\varepsilon_t\|^{2+\epsilon}\,\big|\mathcal{F}_{t-1}\big)<\infty$ a.s. for some $\epsilon>0$ without affecting our subsequent results.

Our functional autoregressive model may be used to study the dynamics of different characteristics of a functional time series. To be specific, for any $v\in H$, we define $\langle v,f_t\rangle$ to be the v-characteristic of f_t, i.e., the characteristic of f_t generated by v. For example, if f_t is the density function of a distribution and v is the k-th order power function defined by $v(x)=x^k$, the v-characteristic of f_t is the k-th moment of the distribution. Now for any $v\in H$, we may consider the process of the v-characteristic given as

$$\langle v,f_t\rangle=\langle v,Af_{t-1}\rangle+\langle v,\varepsilon_t\rangle=\langle A^*v,f_{t-1}\rangle+\varepsilon_t(v),$$

where $(\varepsilon_t(v))$ is a scalar white noise process. We may view $(A^*v)(x)$ as the response of $\langle v,f_t\rangle$ to an impulse to f_{t-1} given by a Dirac-δ function with a spike at x, where the superscript * denotes the adjoint. Similarly, $A^{*i}v$ may be viewed as the response function of $\langle v,f_t\rangle$ to impulses to f_{t-i}.

2.2 Functional Beveridge-Nelson Decomposition

It is very useful to obtain the Beveridge-Nelson decomposition of a functional time series (f_t) generated by an FAR(1) as in (1). To present the functional Beveridge-Nelson decomposition more effectively, we first introduce some notation. In our subsequent discussion, we use the subscripts or superscripts "P" and "T" to denote curves, functions and operators related to the permanent and transitory components of (f_t), respectively. We let Γ_P and Γ_T be two non-intersecting Cauchy contours on the complex plane such that 1 lies in the inner domain of Γ_P and $\lambda(A)\backslash\{1\}$ lies in the inner domain of Γ_T. Such a separation of elements in $\lambda(A)$ is guaranteed, since 1 cannot be a limit point of $\lambda(A)$ by the compactness of A. We define two operators on H by

$$\Pi_P=\frac{1}{2\pi i}\oint_{\Gamma_P}(\lambda-A)^{-1}d\lambda$$

and

$$\Pi_T=\frac{1}{2\pi i}\oint_{\Gamma_T}(\lambda-A)^{-1}d\lambda$$

where the contour integral is defined as the Stieltjes integral and the convergence is in the operator norm. A standard argument in complex analysis shows that the definitions of Π_P and Π_T are independent of the choices of Γ_P and Γ_T. Finally, we denote the images of Π_P

and Π_T respectively by H_P and H_T.

Theorem 2.1. *Let Assumption 2.1 hold. Then we have*
(a) $H = H_P \oplus H_T$,
(b) H_P *and* H_T *are invariant under* A, *and*
(c) H_P *is finite dimensional.*

Part (a) of the above theorem implies that $\Pi_P + \Pi_T = 1$, and we may uniquely decompose

$$f_t = f_t^P + f_t^T, \tag{2}$$

where

$$f_t^P = \Pi_P f_t \quad \text{and} \quad f_t^T = \Pi_T f_t,$$

and similarly, $\varepsilon_t = \varepsilon_t^P + \varepsilon_t^T$ with $\varepsilon_t^P = \Pi_P \varepsilon_t$ and $\varepsilon_t^T = \Pi_T \varepsilon_t$, for $t = 1, 2, \ldots$. Note that here and elsewhere in this paper, we denote the identity operators on H and its subspaces by 1. Part (b) implies that $A f_t^P \in H_P$ and $A f_t^T \in H_T$, and therefore, we may easily deduce that

$$f_t^P = A_P f_{t-1}^P + \varepsilon_t^P \tag{3}$$

and

$$f_t^T = A_T f_{t-1}^T + \varepsilon_t^T \tag{4}$$

for $t = 1, 2, \ldots$, where A_P and A_T denote the restrictions of A on H_P and H_T, respectively. Part (c) means that (f_t^P) is finite dimensional. Figure 1 gives a graphical presentation of our decomposition, where each subspace is represented by a one-dimensional line.

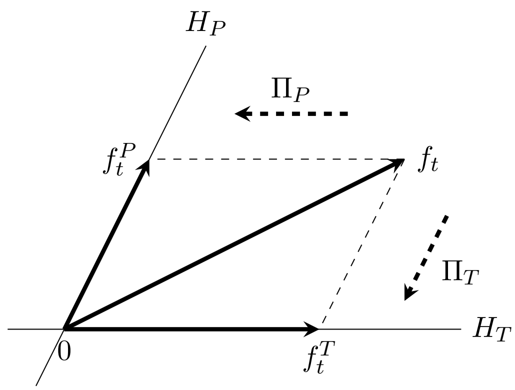
Let H_P be ℓ-dimensional, and A_P be the linear transformation on H_P. It follows that $A_P - 1$ becomes nilpotent of degree d, i.e., d is the smallest integer such that $(A_P - 1)^d = 0$, for some $1 \leqslant d \leqslant \ell$. This is well known. See, e.g., Theorem 2 in Section 58 of Halmos (1974). Furthermore, the degree of nilpotency completely characterizes the order of integration for (f_t^P).

Lemma 2.2. $A_P - 1$ *is nilpotent of degree* d *if and only if* (f_t^P) *is I(d).*

Although processes of higher integrated orders may be useful, time series integrated of order one seems to be most relevant in economic applications. Therefore, we assume that $A_P - 1$ is nilpotent of degree 1, i.e., $A_P = 1$, in which case (f_t^P) becomes a random walk. Moreover, we let $\|A_T^r\| < 1$ for some $r \geqslant 1$, so that (f_t^T) is stationary.

Assumption 2.2. $A_P = 1$ and $\|A_T^r\| < 1$ for some integer $r \geqslant 1$.

Figure 1: Decomposition of Functional Time Series



Notes: This figure illustrates the decomposition of a functional time series (f_t) into its permanent component (f_t^P) and transitory component (f_t^T). The permanent subspace H_P and transitory subspace H_T are represented by one-dimensional lines.

Under Assumptions 2.1 and 2.2, (f_t) becomes an I(1) process with ℓ unit roots. In particular, (3) reduces to

$$f_t^P = f_{t-1}^P + \varepsilon_t^P \tag{5}$$

for $t = 1, 2, \ldots$, and (4) defines a stationary functional autoregressive process (f_t^T). Consequently, we have the following decomposition theorem.

Theorem 2.3. *Let Assumptions 2.1 and 2.2 hold. Then the decomposition introduced in* (2) *becomes the functional Beveridge-Nelson decomposition, with* (f_t^P) *and* (f_t^T) *representing the permanent and transitory components of* (f_t)*, whose dynamics are given by* (5) *and* (4) *respectively.*

It is also useful to introduce the decomposition of the dual space H^* of H corresponding to our decomposition of $H = H_P \oplus H_T$. As is well known, H is its own dual space, i.e., $H = H^*$ by the Riesz representation theorem. We let

$$H^* = H_P^* \oplus H_T^*$$

with

$$H_P^* = H_T^\perp \quad \text{and} \quad H_T^* = H_P^\perp,$$

where $H_P^\perp$ and $H_T^\perp$ are the orthogonal complements of H_P and H_T, respectively.

For $v \in H_P^*$, we may easily deduce that

$$\langle v, f_t \rangle = \langle v, f_t^P \rangle = \langle v, f_{t-1}^P \rangle + \langle v, \varepsilon_t^P \rangle = \langle v, f_{t-1} \rangle + \langle v, \varepsilon_t \rangle.$$

This implies that $(\langle v, f_t \rangle)$ is a random walk. On the other hand, for $v \in H_T^*$, we have

$$\langle v, f_t \rangle = \langle v, f_t^T \rangle,$$

and therefore, $(\langle v, f_t \rangle)$ is a stationary process. In sum, the coordinate process $(\langle v, f_t \rangle)$ becomes a random walk or a stationary process, depending on whether $v \in H_P^*$ or $v \in H_T^*$, respectively.

Let $A^* = A_P^* + A_T^*$, where A_P^* and A_T^* are A^* restricted on H_P^* and H_T^*, respectively.

Lemma 2.4. *Let Assumptions 2.1 and 2.2 hold. Then $A_P^* = 1$.*

Therefore, we have $A_P^* = 1$ as well as $A_P = 1$.

Subsequently, we denote H_P and H_T^* by H_N and H_S, which will be referred to as the nonstationary subspace and the stationary subspace of H, respectively. Under Assumptions 2.1 and 2.2, we have

$$H = H_N \oplus H_S, \tag{6}$$

and for $v \in H_N$ and $v \in H_S$, $(\langle v, f_t \rangle)$ is nonstationary and stationary, respectively. Unlike the decomposition $H = H_P \oplus H_T$ in Theorem 2.1, the decomposition in (6) is orthogonal. We define Π_N and Π_S to be the orthogonal projections on the nonstationary and stationary subspaces H_N and H_S of H, and let

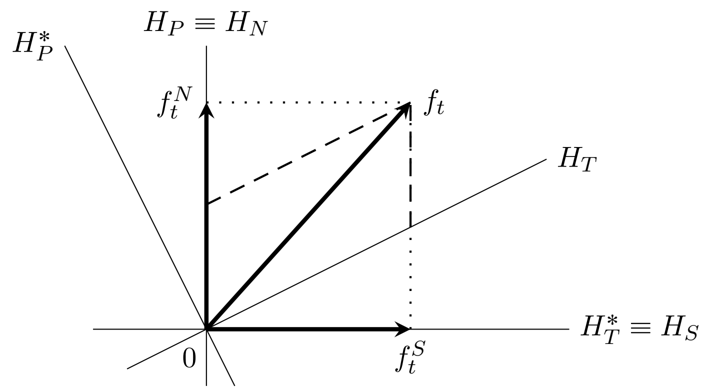
$$f_t^N = \Pi_N f_t \quad \text{and} \quad f_t^S = \Pi_S f_t \tag{7}$$

for $t = 1, 2, \ldots$. Our subsequent theoretical development will rely on the decompositions in (6) and (7). See Figure 2 for the graphical presentation of the decompositions of H and its dual space H^* we introduce. The dotted lines represent the projections Π_N and Π_S, and the dashed lines represent the projections Π_P and Π_T.

2.3 Finite Dimensional Case

To see how our model and framework are related to the existing literature on VAR with unit roots and cointegration, let $H = \mathbb{R}^m$ and (f_t) be a usual m-dimensional time series. In this case, the autoregressive operator A reduces to an $m \times m$ matrix. If we assume that

Figure 2: Decomposed Subspaces



Notes: This figure illustrates the decomposition of the Hilbert space H into the permanent subspace H_P and the transitory space H_T, and the decomposition of the dual space $H^* = H$ into the stationary space H_T^* and the random walk dual space H_P^*. It also presents the decomposition of the functional time series (f_t) into its nonstationary component (f_t^N) and stationary component (f_t^S). The projections on the stationary and nonstationary subspaces are represented by dotted lines, and the projections on the permanent and transitory subspaces are represented by dashed lines.

there are ℓ unit roots for $0 < \ell < m$, we may let

$$A = 1 + \alpha\beta'$$

and write

$$\Delta f_t = \alpha\beta' f_{t-1} + \varepsilon_t, \tag{8}$$

where α and β, which are identified only up to their ranges, are $m \times (m-\ell)$ matrices of parameters such that the $(m-\ell) \times (m-\ell)$ matrix $\alpha'\beta$ is nonsingular. Under Assumptions 2.1 and 2.2, we may write the FAR in (1) as the ECM in (8), where (f_t) is I(1), and $(\beta' f_t)$ is stationary with each column of β representing a cointegrating relationship in (f_t). In our subsequent discussion, we denote by $\alpha_\perp$ and $\beta_\perp$ the $m \times \ell$ matrices such that $\alpha_\perp'\alpha = 0$ and $\beta_\perp'\beta = 0$, where $\alpha_\perp$ and $\beta_\perp$ are again identified only up to their ranges.

For (f_t) generated by the ECM in (8), we have

$$H_P = \mathcal{R}(\beta_\perp) \quad \text{and} \quad H_T = \mathcal{R}(\alpha),$$

since $A\beta_\perp = \beta_\perp$, $A\mathcal{R}(\alpha) \subset \mathcal{R}(\alpha)$, and $\mathcal{R}(\alpha) \oplus \mathcal{R}(\beta_\perp) = \mathbb{R}^m$. Furthermore, it follows that

$$H_P^* = H_T^\perp = \mathcal{R}(\alpha_\perp) \quad \text{and} \quad H_T^* = H_P^\perp = \mathcal{R}(\beta),$$

which implies that $(\alpha_\perp' f_t)$ is a unit root process and $(\beta' f_t)$ is a stationary process. We may explicitly obtain the projections Π_P and Π_T as

$$\Pi_P = \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' \quad \text{and} \quad \Pi_T = \alpha(\beta'\alpha)^{-1}\beta',$$

respectively. Note that the subspace H_P is defined by Granger as the *attractor space*, and the subspace H_T^* is often referred to as the *cointegrating space*.

Recall that we also define H_P and H_T^* to be the nonstationary subspace H_N and the stationary subspace H_S, respectively, which decompose $H = \mathbb{R}^p$ into two orthogonal subspaces. The projections Π_N and Π_S on the two orthogonal subspaces H_N and H_S are given by

$$\Pi_N = \beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp' \quad \text{and} \quad \Pi_S = \beta(\beta'\beta)^{-1}\beta',$$

respectively.

3 Estimation, Prediction and Asymptotic Theory

3.1 Preliminaries

Functional Limit Theory For our asymptotics of FAR with unit roots, we need an invariance principle in H.

Lemma 3.1. *Let Assumptions 2.1 and 2.2 hold. If we define*

$$\overline{W}(r) = \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\varepsilon_t$$

for $r \in [0,1]$, then $\overline{W} \to_d W$ as $n \to \infty$, where W is Brownian motion on H with variance operator Σ.

For more discussions on Brownian motion in Hilbert space, the reader is referred to Kuelbs (1973).

For (f_t^N), we write

$$f_t^N = f_t^P + (f_t^T - f_t^S)$$

for $t = 1, 2, \ldots$, and note that (f_t^P) is a random walk driven by the innovation (ε_t^P), i.e., $f_t^P = f_{t-1}^P + \varepsilon_t^P$ for $t = 1, 2, \ldots$, and that $(f_t^T - f_t^S)$ is stationary. It follows from Lemma 3.1 that if we define

$$\overline{W}_P(r) = \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\varepsilon_t^P$$

for $r \in [0,1]$, then $\overline{W}_P \to_d W_P$ as $n \to \infty$, where W_P is a Brownian motion on H_P with variance $\Sigma_P = \Pi_P \Sigma \Pi_P^*$. Therefore, if properly normalized, (f_t^P) behaves like W_P in the limit. Note that Σ_P is finite dimensional and of rank ℓ, which implies that W_P is degenerate and takes values only in an ℓ-dimensional subspace H_P of H. On the other hand, since $f_t^T = A_T f_{t-1}^T + \varepsilon_t^T$ for $t = 1, 2, \ldots$, the variance operator of (f_t^T) is given by $\sum_{k=0}^{\infty} A_T^k \Sigma_T A_T^{*k}$ with $\Sigma_T = \Pi_T \Sigma \Pi_T^*$. Since

$$f_t^S = \Pi_S f_t^T,$$

the variance operator of (f_t^S) is given by $\Sigma_S = \Pi_S \big(\sum_{k=0}^{\infty} A_T^k \Sigma_T A_T^{*k} \big) \Pi_S$.

Asymptotics for Sample Variance Operator Let n be the sample size, and define the unnormalized sample variance operator $\widehat{\Gamma}$ of (f_t) by

$$\widehat{\Gamma} = \sum_{t=1}^{n} (f_t \otimes f_t), \tag{9}$$

which is decomposed as

$$\widehat{\Gamma} = n^2 \overline{\Gamma}_{NN} + n \overline{\Gamma}_{NS} + n \overline{\Gamma}_{SN} + n \overline{\Gamma}_{SS}, \tag{10}$$

where

$$\overline{\Gamma}_{NN} = \frac{\Pi_N \widehat{\Gamma} \Pi_N}{n^2} = \frac{1}{n^2} \sum_{t=1}^{n} (f_t^N \otimes f_t^N),$$

$$\overline{\Gamma}_{SS} = \frac{\Pi_S \widehat{\Gamma} \Pi_S}{n} = \frac{1}{n} \sum_{t=1}^{n} (f_t^S \otimes f_t^S),$$

$$\overline{\Gamma}_{NS} = \frac{\Pi_N \widehat{\Gamma} \Pi_S}{n} = \frac{1}{n} \sum_{t=1}^{n} (f_t^N \otimes f_t^S),$$

and $\overline{\Gamma}_{SN} = \overline{\Gamma}_{NS}^*$.

Lemma 3.2. *Let Assumptions 2.1 and 2.2 hold. Then*

$$\overline{\Gamma}_{NN} \to_d \Gamma_{NN} = \int_0^1 (W_P \otimes W_P)(r) dr,$$

$$\overline{\Gamma}_{SS} \to_p \Gamma_{SS} = \Pi_S \left(\sum_{k=0}^{\infty} A_T^k \Sigma_T A_T^{*k} \right) \Pi_S$$

as $n \to \infty$. Moreover,

$$\overline{\Gamma}_{NS} = O_p(1) \quad \textit{and} \quad \overline{\Gamma}_{SN} = O_p(1)$$

for large n.

On the ℓ-dimensional nonstationary subspace H_N of H, the sample variance operator of (f_t), if normalized by n^{-2}, converges in distribution to Γ_{NN}, which is a random operator. On the other hand, on the stationary space H_S of H, the usual sample variance operator of (f_t) normalized by n^{-1} converges in probability to its common variance operator. In fact, it follows from Hu et al. (2016) that

$$\left\|\overline{\Gamma}_{SS} - \Gamma_{SS}\right\| = O\left(n^{-1/2}\log^{1/2} n\right) \quad \text{a.s.,}$$

under Assumptions 2.1 and 2.2. Note that

$$\left(\frac{\Pi_N}{n} + \frac{\Pi_S}{\sqrt{n}}\right)\widehat{\Gamma}\left(\frac{\Pi_N}{n} + \frac{\Pi_S}{\sqrt{n}}\right) = \frac{\overline{\Gamma}_{NN}}{n^2} + \frac{\overline{\Gamma}_{SS}}{n} + O_p(n^{-1/2}),$$

which implies that, if we normalize Π_N and Π_S appropriately, the terms $\overline{\Gamma}_{NS}$ and $\overline{\Gamma}_{SN}$ become negligible in the limit and do not appear in our asymptotics. The unit root and stationary components of (f_t) are therefore asymptotically orthogonal. This extends the asymptotic orthogonality of the unit root and stationary components in finite dimensional nonstationary time series, which is shown in, e.g., Park and Phillips (1989).

Functional Principal Component Analysis Since $\widehat{\Gamma}$ is self-adjoint and positive semi-definite, it has real and nonnegative eigenvalues, $\hat{\lambda}_1 \geqslant \cdots \geqslant \hat{\lambda}_n$, with the corresponding eigenfunctions $\hat{v}_1, \ldots, \hat{v}_n$, which are orthogonal. We may assume that $\|\hat{v}_k\| = 1$ for $k = 1, \ldots, n$. In fact, the eigenfunctions $\hat{v}_1, \ldots, \hat{v}_n$ are the (normalized) functional principal components which are used widely in functional data analysis. We let $(\lambda_k(\Gamma_{NN}), v_k(\Gamma_{NN}))$ be the pairs of eigenvalues and eigenfunctions of Γ_{NN} such that $\lambda_k(\Gamma_{NN})$'s are in descending order. Similarly, we define $(\lambda_k(\Gamma_{SS}), v_k(\Gamma_{SS}))$ to be the ordered pairs of eigenvalues and eigenfunctions of Γ_{SS} such that $\lambda_k(\Gamma_{SS})$'s are in descending order. For expositional convenience, we assume that the eigenvalues $(\lambda_k)_{k>\ell}$ are different from each other.[3] For definiteness, we assume that $(v_k(\Gamma_{NN}))$ and $(v_k(\Gamma_{SS}))$ are normalized and that their signs are aligned with $(\hat{v}_k)$, i.e., $\langle \hat{v}_k, v_k(\Gamma_{NN})\rangle \geqslant 0$ for $k = 1, \ldots, \ell$, and $\langle \hat{v}_{k+\ell}, v_k(\Gamma_{SS})\rangle \geqslant 0$ for $k = 1, 2, \ldots$. The following lemma follows directly from Chang et al. (2016b).

[3] We could potentially allow for multiplicity. However in that case the eigenfunctions could not be uniquely identified even after normalization, which introduces expositional complications.

Lemma 3.3. *Let Assumptions 2.1 and 2.2 hold. Then*

$$\left(n^{-2}\hat{\lambda}_k, \hat{v}_k\right) \to_d (\lambda_k(\Gamma_{NN}), v_k(\Gamma_{NN}))$$

as $n \to \infty$ jointly for $k = 1, \ldots, \ell$, and

$$\left(n^{-1}\hat{\lambda}_{k+\ell}, \hat{v}_{k+\ell}\right) \to_p (\lambda_k(\Gamma_{SS}), v_k(\Gamma_{SS}))$$

as $n \to \infty$ for $k = 1, 2, \ldots$.

The eigenvalues associated with the nonstationary and stationary subspaces diverge at different rates, and this was used by Chang et al. (2016b) to develop a consistent test for the number of unit roots, or equivalently, the dimension of nonstationary subspace ℓ in general functional time series with unit roots. The ℓ leading functional principal components $(\hat{v}_k)_{k=1}^{\ell}$ converge in distribution to the ordered eigenfunctions $\left(v_k(\Gamma_{NN})\right)_{k=1}^{\ell}$ of the random operator Γ_{NN} on H_N introduced in Lemma 3.2. Although $\left(v_k(\Gamma_{NN})\right)_{k=1}^{\ell}$ span H_N a.s., they are not deterministic but random functions. The rest functional principal components $(\hat{v}_{k+\ell})_{k=1}^{n-\ell}$ converge in probability to the ordered eigenfunctions $\left(v_k(\Gamma_{SS})\right)_{k=1}^{n-\ell}$ of the deterministic operator Γ_{SS} defined explicitly in Lemma 3.2. It is clear that we need normalizations by n^{-2} and n^{-1} for the eigenvalues $\left(\hat{\lambda}_k\right)_{k=1}^{\ell}$ and $\left(\hat{\lambda}_{k+\ell}\right)_{k=1}^{n-\ell}$, respectively, by Lemma 3.2. Throughout the paper, we let $v_{k+\ell} = v_k(\Gamma_{SS})$ for $k = 1, 2, \ldots$ for notational brevity, so that $(v_{k+\ell})_{k=1}^{\infty}$ spans H_S. Moreover, we let $(v_k)_{k=1}^{\ell}$ be an arbitrary set of functions, which spans H_N. Then $(v_k)_{k=1}^{\infty}$ is an orthonormal basis of H such that $(v_k)_{k=1}^{\ell}$ spans H_N and $(v_k)_{k=\ell+1}^{\infty}$ spans H_S.

Once $(\hat{v}_k)$ are obtained and the number ℓ of unit roots is known, we may estimate the projection Π_N on the nonstationary subspace H_N by

$$\widehat{\Pi}_N = \sum_{k=1}^{\ell} (\hat{v}_k \otimes \hat{v}_k), \tag{11}$$

and the projection Π_S on the stationary subspace $H_S = H_N^{\perp}$ by $\widetilde{\Pi}_S = 1 - \widehat{\Pi}_N$. As shown in Chang et al. (2016b), we have

$$\widehat{\Pi}_N = \Pi_N + O_p(n^{-1}) \quad \text{and} \quad \widetilde{\Pi}_S = \Pi_S + O_p(n^{-1})$$

for large n, under Assumptions 2.1 and 2.2. As discussed, we may consistently estimate the nonstationary space H_N by the subspace of H spanned by $(\hat{v}_k)_{k=1}^{\ell}$, although $(\hat{v}_k)_{k=1}^{\ell}$ does not converge to $(v_k)_{k=1}^{\ell}$ spanning H_N a.s. or in probability.

Let

$$\widehat{f}_t^N = \widehat{\Pi}_N f_t \quad \text{and} \quad \widetilde{f}_t^S = \widetilde{\Pi}_S f_t, \tag{12}$$

and redefine $\overline{\Gamma}_{NN}, \overline{\Gamma}_{SS}, \overline{\Gamma}_{NS}$ and $\overline{\Gamma}_{SN}$ using $(\widehat{f}_t^N, \widetilde{f}_t^S)$ in place of (f_t^N, f_t^S), respectively. Then the differences between the newly defined $\overline{\Gamma}_{NN}, \overline{\Gamma}_{SS}, \overline{\Gamma}_{NS}$ and their original versions are only of order $O_p(n^{-1})$. In particular, the newly defined $\overline{\Gamma}_{NN}$ and $\overline{\Gamma}_{SS}$ are asymptotically equivalent to their original versions, whose asymptotics are derived in Lemma 3.2. Therefore, we will not distinguish the new versions from the old ones.

Ill-Posed Inverse Problem We may easily see that Γ_{NN} is invertible on H_N and Γ_{NN}^+ is well defined. However, $\Gamma_{SS} = \mathbb{E}(f_t^S \otimes f_t^S)$ is not invertible on H_S, since $\Gamma_{SS} = \sum_{k=\ell+1}^{\infty} \lambda_k (v_k \otimes v_k)$ with $\sum_{k=\ell+1}^{\infty} \lambda_k < \infty$ under our condition $\mathbb{E}\|f_t^S\|^2 < \infty$, which means in particular that $\lambda_k \to 0$ as $k \to \infty$, and therefore, Γ_{SS}^+ is not well defined.[4] This creates the so-called ill-posed inverse problem in estimating the autoregressive operator A, which involves inversion of the sample variance operator of (f_t). To deal with this ill-posed inverse problem, we use the standard approach in functional data analysis, which will be explained below.

Let m_n be a sequence of numbers such that $\ell < m_n < n$ and $m_n \to \infty$ with $m_n/n \to 0$ as $n \to \infty$, which we subsequently write m instead of m_n for notational brevity. Moreover, let

$$\widehat{f}_t = \widehat{\Pi} f_t \quad \text{and} \quad \widehat{f}_t^S = \widehat{\Pi}_S f_t \tag{13}$$

for $t = 1, \ldots, n$, where

$$\widehat{\Pi} = \sum_{k=1}^{m} (\hat{v}_k \otimes \hat{v}_k), \quad \text{and} \quad \widehat{\Pi}_S = \sum_{k=\ell+1}^{m} (\hat{v}_k \otimes \hat{v}_k), \tag{14}$$

from which it follows immediately that

$$\widehat{f}_t = \widehat{f}_t^N + \widehat{f}_t^S \quad \text{and} \quad \widehat{\Pi} = \widehat{\Pi}_N + \widehat{\Pi}_S$$

for $t = 1, \ldots, n$, where $\widehat{\Pi}_N$ and $\widehat{f}_t^N$ are defined in (11) and (12), respectively. Note that $\widehat{\Pi}_N + \widehat{\Pi}_S \neq 1$, which is in contrast with $\widehat{\Pi}_N + \widetilde{\Pi}_S = 1$, where $\widetilde{\Pi}_S = 1 - \widehat{\Pi}_N$ as defined earlier.

To deal with the ill-posed inverse problem in estimating the autoregressive operator A, we use $(\widehat{f}_t)$ in place of (f_t) to approximate the inverse of the sample variance operator of

[4] In fact, the operator Γ_{SS} with $\sum_{k=\ell+1}^{\infty} \lambda_k < \infty$ is said to be nuclear or trace-class.

(f_t). Note that

$$\left(\sum_{t=1}^{n}(f_t \otimes f_t)\right)^+ = \sum_{k=1}^{n} \frac{1}{\hat{\lambda}_k}(\hat{v}_k \otimes \hat{v}_k),$$

while

$$\left(\sum_{t=1}^{n}(\widehat{f}_t \otimes \widehat{f}_t)\right)^+ = \sum_{k=1}^{m} \frac{1}{\hat{\lambda}_k}(\hat{v}_k \otimes \hat{v}_k),$$

which explains the reason why we use $(\widehat{f}_t)$ to solve the ill-posed inverse problem in the sample variance operator of (f_t).

3.2 Estimators and Their Asymptotics

Let

$$\widehat{A} = \left(\sum_{t=1}^{n}(f_t \otimes f_{t-1})\right)\left(\sum_{t=1}^{n}(\widehat{f}_{t-1} \otimes \widehat{f}_{t-1})\right)^+, \tag{15}$$

where $(\widehat{f}_t)$ is defined in (13). This is the commonly used estimator for the autoregressive operator A. Bosq (2000) and Hu et al. (2016) use the same estimator to analyze stationary functional autoregressions.

To develop asymptotics for $\widehat{A}$, we define a sequence (τ_k) for $k = \ell+1, \ell+2, \ldots$ by $\tau_{\ell+1} = 2\sqrt{2}(\lambda_{\ell+1} - \lambda_{\ell+2})^{-1}$ and $\tau_k = 2\sqrt{2}\max\{(\lambda_{k-1} - \lambda_k)^{-1}, (\lambda_k - \lambda_{k+1})^{-1}\}$ for $k > \ell+1$, and introduce the following assumption.

Assumption 3.1. $\log n\big(\sum_{k=\ell+1}^{m}\tau_k\big)^2/\big(n\lambda_m^2\big) \to 0$ as $n \to \infty$.

Note that Assumption 3.1 does not put any actual restrictions on the time series (f_t) itself. Since $\sum_{k=\ell+1}^{m}\tau_k$ is increasing in m and λ_m is decreasing in m, it merely controls how fast m may grow as $n \to \infty$. That is, it only imposes a restriction on how we may choose m as a function of n. The following theorem provides asymptotics for our autoregressive operator estimator $\widehat{A}$. In what follows, we let

$$\overline{A} = A\widehat{\Pi},$$

where $\widehat{\Pi}$ is defined earlier in (14).

Theorem 3.4. *Let Assumptions 2.1, 2.2 and 3.1 hold. Then*

$$\left\|\widehat{A} - A\right\| \to_p 0$$

as $n \to \infty$. On H_N, we have

$$n(\widehat{A} - \overline{A}) \to_d \left(\int_0^1 (dW \otimes W_P)\right)\left(\int_0^1 (W_P \otimes W_P)\right)^+$$

as $n \to \infty$. Moreover, for any $v \notin H_N$, we have

$$\frac{\sqrt{n}}{s_m(v)}(\widehat{A} - \overline{A})v \to_d \mathbb{N}(0, \Sigma)$$

as $n \to \infty$, where $s_m^2(v) = \sum_{k=\ell+1}^m \lambda_k^{-1}\langle v_k, v\rangle^2$, and $\mathbb{N}(0, \Sigma)$ is a Gaussian random element taking values in H with mean zero and variance operator Σ.

Our estimator $\widehat{A}$ of the autoregressive operator A in (15) is consistent. The limit behaviors of $\widehat{A}$ are quite distinctive on H_N and elsewhere. In fact, $\widehat{A}$ requires distinctive normalization factors and yields different types of limit distributions on H_N and elsewhere. On H_N, $n(\widehat{A} - \overline{A})$ converges weakly in operator norm to a nonstandard distribution represented as a function of Brownian motions. Unfortunately, as shown in Mas (2007), $\widehat{A} - \overline{A}$ does not converge weakly in any norm topology under any normalization. Nevertheless, we may still consider pointwise weak convergence to establish asymptotic normality of $(\widehat{A} - \overline{A})v$ for $v \notin H_N$. Note that the convergence rate of $(\widehat{A} - \overline{A})v$ for $v \notin H_N$ depends on v. Specifically, $(\widehat{A} - \overline{A})v$ converges at the usual $\sqrt{n}$-rate if $\sum_{k=1}^\infty \lambda_k^{-1}\langle v_k, v\rangle^2 < \infty$, and converges at a rate slower than $\sqrt{n}$ if $\sum_{k=1}^\infty \lambda_k^{-1}\langle v_k, v\rangle^2 = \infty$.

Our nonstationary asymptotics in Theorem 3.4 are coordinate-free, and do not rely on any particular coordinate system. When $H = \mathbb{R}^m$ and $H_N = \mathcal{R}(\beta_\perp)$ as in Section 2.3, we may present our asymptotics more explicitly using a coordinate system given by the column vectors of a particular choice of $\beta_\perp$, in which we represent $v \in H_N$ as $u \in \mathbb{R}^\ell$ such that $v = \beta_\perp u$.[5] If such a coordinate system is used, we have

$$n(\widehat{A} - 1)\beta_\perp \to_d \left(\int_0^1 dW V_P'\right)\left(\int_0^1 V_P V_P'\right)^{-1}, \tag{16}$$

where V_P is an ℓ-dimensional Brownian motion defined as

$$W_P = \beta_\perp V_P \quad \text{with} \quad V_P = (\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' W.$$

Note that $\overline{A} = A$ in this case, and V_P is uniquely defined with any choice of $\alpha_\perp$ such

[5]Here we assume that $\beta_\perp$ is of full column rank. As is well known, $\mathcal{R}(\beta_\perp) = \mathcal{R}(\beta_\perp T)$ for any nonsingular matrix T, and therefore, $\beta_\perp$ is not identified uniquely for a given H_N. Therefore, we must choose a particular $\beta_\perp$ for which $(\beta_\perp) = H_N$ to use it as a coordinate system here.

that $\mathcal{R}(\alpha_\perp) = H_P^*$. The asymptotics in (16) may be viewed as a coordinate version of our asymptotics in Theorem 3.4. As expected, if $m = \ell = 1$, our limit distribution reduces to a scaled Dickey-Fuller distribution.

Theorem 3.4 shows that our estimator $\widehat{A}$ of the autoregressive operator A contains bias terms on both the nonstationarity and stationarity subspaces, i.e., H_N and H_S. To analyze the bias terms, it is necessary to introduce some technical conditions.

Assumption 3.2. We assume that

(a) (λ_k) is convex in k for k large enough,

(b) $n^{-1/2}m^{5/2}\log^2 m \to 0$, and

(c) $\sum_{i=\ell+1}^{m}\sum_{j=m+\ell+1}^{\infty}\lambda_i\lambda_j/(\lambda_i-\lambda_j)^2 = o(m)$

as $n \to \infty$.

The condition in (a) is mild and is satisfied by many sequences of eigenvalues decaying at polynomial and exponential rates. The condition in (b) holds as long as $m = O(n^{1/5-\delta})$ for any $\delta > 0$, and m does not grow too fast as $n \to \infty$. The condition in (c) is more stringent, though not prohibitively so. For many practical applications, it appears that (λ_k) decays geometrically and we may set $\lambda_k = \rho^k$ for some $0 < \rho < 1$. In this case, we may easily deduce that $\sum_{i=\ell+1}^{m}\sum_{j=m+\ell+1}^{\infty}\lambda_i\lambda_j/(\lambda_i-\lambda_j)^2 = O(1)$.

Let $\underline{H}_S$ be the subspace of H spanned by $(v_k)_{k=\ell+1}^m$ and define $\underline{\Pi}_S$ to be the orthogonal projection on $\underline{H}_S$, and similarly, let $\underline{H}$ be the subspace of H spanned by $(v_k)_{k=1}^m$ and define $\underline{\Pi}$ to be the orthogonal projection on $\underline{H}$. Note that $\underline{H}_S$ is an m-dimensional subspace of H_S, and that $\underline{H} = H_N \oplus \underline{H}_S$ and $\underline{\Pi} = \Pi_N + \underline{\Pi}_S$.

Corollary 3.5. *Let Assumptions 2.1, 2.2 and 3.2 hold. Then for any $v \in H_N$,*

$$(\overline{A} - A)v = o_p(n^{-1/2}m^{1/2}),$$

and, for any $v \notin H_N$,

$$(\overline{A} - A)v = o_p(n^{-1/2}m^{1/2}) + O\left(\|A(1-\underline{\Pi})v\|\right)$$

for large n.

Corollary 3.5 provides the orders of the bias terms in our autoregressive operator estimator $\widehat{A}$. The bias terms become negligible as long as $m \to \infty$ as $n \to \infty$. Note that $(1-\underline{\Pi})v \to 0$ as $m \to \infty$ for any $v \in H$.

Once we obtain the autoregressive operator estimator, we may obtain the residuals by

$$\widehat{\varepsilon}_t = f_t - \widehat{A}f_{t-1}$$

and estimate Σ by

$$\widehat{\Sigma} = \frac{1}{n} \sum_{t=1}^{n} (\widehat{\varepsilon}_t \otimes \widehat{\varepsilon}_t).$$

The following corollary is an obvious consequence of Theorem 3.4.

Corollary 3.6. *Let Assumptions 2.1, 2.2 and 3.1 hold. Then*

$$\left\| \widehat{\Sigma} - \Sigma \right\| \to_p 0$$

as $n \to \infty$.

In obtaining the estimator $\widehat{A}$ of A introduced earlier in (15), we do not impose the restrictions implied by the presence of unit roots in (f_t). In the following, we propose another estimator $\widetilde{A}$ of A with those restrictions, which is defined as

$$\widetilde{A} = \widehat{\Pi} + \left(\sum_{t=1}^{n} (\Delta f_t \otimes f_{t-1}) \right) \left(\sum_{t=1}^{n} \left(\widehat{f}_{t-1}^S \otimes \widehat{f}_{t-1}^S \right) \right)^+, \tag{17}$$

and let

$$\widetilde{B} = (\widetilde{A} - 1)\widehat{\Pi} \quad \text{and} \quad \overline{B}_S = (A - 1)\widehat{\Pi}_S.$$

Note in particular that $\widetilde{B}v = v$ and $\overline{B}_S v = 0$ for any $v \in \widehat{H}_N$. We may easily deduce that

Theorem 3.7. *Let Assumptions 2.1, 2.2 and 3.1 hold. Then*

$$\left\| \widetilde{A} - A \right\| \to_p 0$$

as $n \to \infty$. On H_N, we have

$$\left\| \widetilde{B} - \overline{B}_S \right\| = o_p(n^{-1})$$

for large n. Moreover, for any $v \notin H_N$, we have

$$\frac{\sqrt{n}}{s_m(v)} \left(\widetilde{B} - \overline{B}_S \right) v \to_d \mathbb{N}(0, \Sigma)$$

as $n \to \infty$, where $s_m^2(v) = \sum_{k=\ell+1}^{m} \lambda_k^{-1} \langle v_k, v \rangle^2$, and $\mathbb{N}(0, \Sigma)$ is a Gaussian random element taking values in H with mean zero and variance operator Σ.

3.3 Beveridge-Nelson Decomposition

To estimate the Beveridge-Nelson decomposition, we need consistent estimators of Π_P and Π_T, i.e., the (non-orthogonal) projections on H_P and H_T along H_T and H_P, respectively. We define their estimators as

$$\widehat{\Pi}_P = \widehat{\Pi}_N - \left(\sum_{t=1}^{n}(\Delta \widehat{f}_t^N \otimes \widehat{f}_{t-1}^S)\right)\left(\sum_{t=1}^{n}(\Delta \widehat{f}_t^S \otimes \widehat{f}_{t-1}^S)\right)^+$$

$$\widehat{\Pi}_T = \widehat{\Pi}_S + \left(\sum_{t=1}^{n}(\Delta \widehat{f}_t^N \otimes \widehat{f}_{t-1}^S)\right)\left(\sum_{t=1}^{n}(\Delta \widehat{f}_t^S \otimes \widehat{f}_{t-1}^S)\right)^+$$

to be our estimators for Π_P and Π_T, respectively. Instead of $\widehat{\Pi}_T$, we may also use

$$\begin{aligned}\widetilde{\Pi}_T &= \widetilde{\Pi}_S + \left(\sum_{t=1}^{n}(\Delta \widehat{f}_t^N \otimes \widehat{f}_{t-1}^S)\right)\left(\sum_{t=1}^{n}(\Delta \widehat{f}_t^S \otimes \widehat{f}_{t-1}^S)\right)^+ \\ &= \left(\sum_{t=1}^{n}(\Delta f_t \otimes \widehat{f}_{t-1}^S)\right)\left(\sum_{t=1}^{n}(\Delta \widehat{f}_t^S \otimes \widehat{f}_{t-1}^S)\right)^+\end{aligned}$$

as an estimator for Π_T. Note that $\widehat{\Pi}_P + \widetilde{\Pi}_T = 1$, whereas $\widehat{\Pi}_P + \widehat{\Pi}_T \neq 1$.

Theorem 3.8. *Let Assumptions 2.1, 2.2, 3.1 and 3.2 hold. Then*

$$\left\|\widehat{\Pi}_P - \Pi_P\right\| \to_p 0 \quad \text{and} \quad \left\|\widetilde{\Pi}_T - \Pi_T\right\| \to_p 0$$

as $n \to \infty$.

Theorem 3.8 shows that the (non-orthogonal) projections Π_P and Π_T along H_T and H_P, respectively, can be consistently estimated by $\widehat{\Pi}_P$ and $\widetilde{\Pi}_T$ in operator norm. On the other hand, $\|\widehat{\Pi}_T - \Pi_T\| \not\to_p 0$.[6] Nevertheless, we have $\widehat{\Pi}_T \to_p \Pi_T$ pointwise as shown below.

Corollary 3.9. *Let Assumptions 2.1, 2.2, 3.1 and 3.2 hold. Then, for any* $v \in H$,

$$\big(\widehat{\Pi}_T - \Pi_T\big)v \to_p 0$$

as $n \to \infty$.

We may also easily deduce that $\big(\widehat{\Pi}_T - \widetilde{\Pi}_T\big)v \to_p 0$ for any $v \in H$, since $\big(\widehat{\Pi}_T - \widetilde{\Pi}_T\big)v = \big(\widehat{\Pi}_T - \Pi_T\big)v - \big(\widetilde{\Pi}_T - \Pi_T\big)v$ and convergence in operator norm implies pointwise convergence.

The estimated Beveridge-Nelson decomposition may now be obtained using $\widehat{\Pi}_P$ and $\widehat{\Pi}_T$, i.e., $f_t = \widehat{f}_t^P + \widehat{f}_t^T$, where

$$\widehat{f}_t^P = \widehat{\Pi}_P f_t \quad \text{and} \quad \widehat{f}_t^T = \widehat{\Pi}_T f_t.$$

[6] This is because $\widehat{\Pi}_T$ is of finite rank and it cannot converge to a non-compact operator Π_T in operator norm.

Clearly, we may also use $\widetilde{\Pi}_T$ in place of $\widehat{\Pi}_T$ to define $\widetilde{f}_t^T$, say, for $t = 1, \ldots, n$. It also follows from Theorem 3.8 that

$$\begin{aligned}
\Pi_P &= \Pi_N - \Big(\mathbb{E}(\Delta f_t^N \otimes f_{t-1}^S)\Big)\Big(\mathbb{E}(\Delta f_t^S \otimes f_{t-1}^S)\Big)^+ \\
\Pi_T &= \Pi_S + \Big(\mathbb{E}(\Delta f_t^N \otimes f_{t-1}^S)\Big)\Big(\mathbb{E}(\Delta f_t^S \otimes f_{t-1}^S)\Big)^+ \\
&= \Big(\mathbb{E}(\Delta f_t \otimes f_{t-1}^S)\Big)\Big(\mathbb{E}(\Delta f_t^S \otimes f_{t-1}^S)\Big)^+,
\end{aligned}$$

which define Π_P and Π_T in terms of Π_N, Π_S and various product moments of (f_t) and (Δf_t).

3.4 Forecast

Our model can be used to make forecasts. We may obtain the one-step forecast as

$$\widehat{f}_{n+1} = \widehat{A} f_n,$$

where $\widehat{A}$ is the estimated autoregressive operator defined in (15). Multiple-step forecasts may be obtained by recursive one-step forecasts. The following results give the asymptotic normality of the predictor. As one would see in the proof of the following lemma, in the prediction procedure we follow Mas (2007) to compute $\widehat{A}$ using data only up to time $n-1$ to avoid technicalities.

Assumption 3.3. We assume that

(a) $\left\|\Gamma_{SS}^{-1/2} A\right\| < \infty$, and

(b) $\sup_{k>\ell} \mathbb{E}\langle v_k, f_t^S\rangle^4/\lambda_k^2 < K$ for some constant K.

Loosely put, condition (a) requires that A be at least as smooth as $\Gamma_{SS}^{1/2}$ on H_S. Condition (b) is satisfied whenever the tail probability of $\langle v_k, f_t^S\rangle$ decreases fast enough. For example, when (f_t^S) is Gaussian, condition (b) holds with $K = 3$.

Lemma 3.10. *Let Assumptions 2.1, 2.2, 3.1 and 3.3 hold. Then*

$$\sqrt{n/m}(\widehat{A} - \overline{A}) f_n \to_d \mathbb{N}(0, \Sigma)$$

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

Once again there is a bias term $\overline{A} f_n - A f_n$ in the result above. To get rid of the bias term so as to obtain the confidence interval for $\hat{f}_{n+1}$, we need Assumption 3.2 as well as an additional assumption.

Assumption 3.4. $(n/m)\sum_{k=m+1}^{\infty}\lambda_k \to 0$ as $n \to \infty$.

For geometrically decaying sequence of eigenvalues $\lambda_k = \rho^k$, we may show easily that $\sum_{k=m+1}^{\infty}\lambda_k = O(\rho^m)$. Therefore, we may set m such that $n = \rho^{-m}$.

Theorem 3.11. *Let Assumptions 2.1, 2.2, 3.1, 3.2, 3.3 and 3.4 hold. Then*

$$\sqrt{n/m}(\widehat{A} - A)f_n \to_d \mathbb{N}(0, \Sigma)$$

as $n \to \infty$.

From Theorem 3.11 we may easily deduce that for any Gaussian (ε_t), we have that

$$\widehat{f}_{n+1} - f_{n+1} = (\widehat{A}_n - A)f_n - \varepsilon_{n+1} \approx_d \mathbb{N}\left(0, \left(1 + \frac{m}{n}\Sigma\right)\right).$$

Consequently, for any $v \in H$, the α-level confidence interval for the forecast of $\langle v, f_n\rangle$ is

$$\left[\langle v, \widehat{f}_{n+1}\rangle - z_{\alpha/2}\sqrt{(1 + m/n)\langle v, \Sigma v\rangle}, \langle v, \widehat{f}_{n+1}\rangle + z_{\alpha/2}\sqrt{(1 + m/n)\langle v, \Sigma v\rangle}\right] \tag{18}$$

where $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ and Φ is the cumulative distribution function of the standard normal distribution.

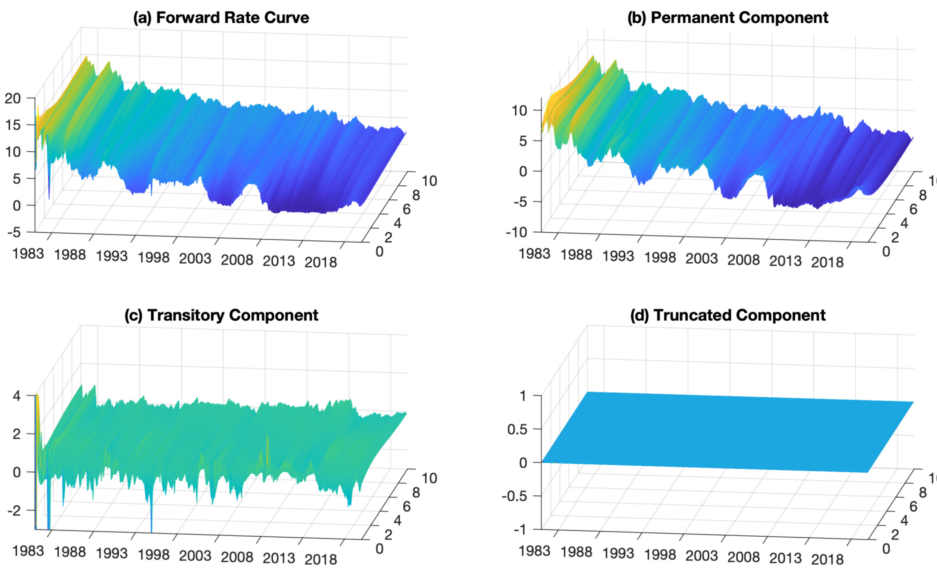
4 Term Structure of US Government Bond Yields

4.1 Preliminaries

The term structure of interest rates is one of the most important topics in finance and macroeconomics. Since US government bonds carry almost no risk, their interest rates are usually viewed as benchmark interest rates.

Let $P_t(\tau)$ be the price of a discount bond at time t that promises to pay \$1 τ years ahead. The yield to maturity $y_t(\tau)$ at time t is defined as the average rate of return of holding this bond until maturity, where τ is the time to maturity. Under continuous compounding, $P_t(\tau) = \exp(-\tau y_t(\tau))$. The yield to maturity $y_t(\tau)$ thus can be calculated as $y_t(\tau) = -\frac{1}{\tau}\ln P_t(\tau)$ once we observe the price of the discount bond. The graph of y_t as a function of the time to maturity τ is called the yield curve at time t. The instantaneous spot rate, denoted by r_t, is the limit of $y_t(\tau)$ as τ approaches zero. In the continuous compounding framework, it measures the current risk-free interest rate.

Figure 3: Time Series of Forward Rate Curves and Its Decompositions



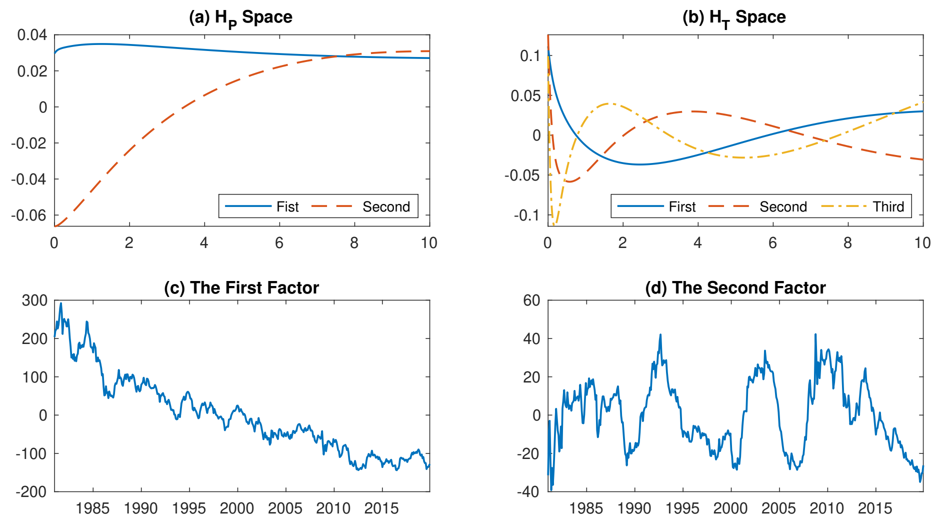
Notes: This figure plots the time series of the end-of-month US government bond forward rate curves from January 1981 to December 2017 and its components. Panel (a) gives the original time series of forward rate curves. Panel (b) and (c) plot its permanent and transitory components, respectively. Panel (d) plots everything that is not in the first five principal components.

The ratio of change in the bond's price at any future time $t + \tau$ defined by $f_t(\tau) = -P_t'(\tau)/P_t(\tau)$ is called the (instantaneous) forward rate, which gives the implied (instantaneous) rate of return of holding the bond at time $t + \tau$ under the no-arbitrage condition. The graph of f_t as a function of the time to maturity τ is called the forward rate curve. The yield curve and the forward rate curve are related through $y_t(\tau) = \frac{1}{\tau}\int_0^\tau f_t(s)ds$. Since the yield curve and the forward rate curve imply each other, they contain the same information. However, it is usually more instructive to loot at the forward rates since they reflect expectations for future interest rates in a more direct way. In this section, we shall study the dynamics of the forward rate curves of US government bonds.

However, the forward rate curves are not directly observable. The Treasury only issues bonds with a limited number of maturities. Gürkaynak et al. (2007) estimate the US Treasury bond forward rate curves using a model of the functional form $f_t(\tau) = \beta_{0t} + \beta_{1t}e^{-\tau/\gamma_{1t}} + \beta_{2t}\frac{\tau}{\gamma_{1t}}e^{-\tau/\gamma_{1t}} + \beta_{3t}\frac{\tau}{\gamma_{2t}}e^{-\tau/\gamma_{2t}}$, where $\beta_{0t}, \beta_{1t}, \beta_{2t}, \beta_{3t}, \gamma_{1t}$ and γ_{2t} are the parameters to be estimated in each period. They estimate the forward rate curves at daily frequency from 1961 on.[7] We use their estimated end-of-month forward rate curves from January

[7]The Federal Reserve Board maintains a web page which posts the update of the estimated forward rate

Figure 4: Factors and Factor Spaces



Notes: Panel (a) plots the first two principal eigenfunctions of the sample variance operator $\widehat{\Gamma}$. These two functions span H_P. Panel (b) plots the first three factor loadings in the space H_T. Panels (c) and (d) present the estimated time series of the level factor and the spread factor. They are estimated as the series of the first and the second principal scores, respectively.

1981 to December 2019. Figure 3(a) plots the time series of the forward rate curves.

There is strong nonstationarity in the forward rate curve process. In general, the trend can be deterministic, stochastic, or a mixture of the two. However, since a deterministic trend suggests predictability, the efficient market hypothesis implies there should be no deterministic trend, or the deterministic trend should be very weak. In this paper, we assume that the trend is stochastic, and use FAR(1) to model the demeaned forward rate curve process.

The functional unit root test developed in Chang et al. (2016b) suggests two unit roots in the demeaned forward rate curve process. Therefore we set $\hat{\ell} = 2$. In addition, we set $m = 5$ to obtain the best rolling out-of-sample forecast performance, in which we use the last one-fifth of periods as the prediction periods.

It turns out that the first five principal components explain 99.98% of variations in the data, which justifies our choice of the value of m. The first two principal components, which correspond to the nonstationary components in the forward rate curve process, explain 99.46% of the data variation.

Panel (a) of Figure 4 presents the first two principal eigenfunctions that span the permanent space H_P. It turns out that the first eigenfunction is very close to the constant

curves quarterly. See http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html.

function. Actually, $\|P_\iota \hat{v}_1\| = 0.996$, where P_ι is the orthogonal projection onto the space spanned by the constant function. Following the conventional terminologies of factor analysis, we shall call the first eigenfunction the loading corresponding to the level factor of the forward rate curves. The second principal eigenfunction is monotonically upward-sloping. Since this component reveals information about the differences between the short rates and the long rates, we shall call it the loading corresponding to the spread factor. The corresponding factors are defined as the inner products of the forward rate curves with the two eigenfunctions, respectively. Panels (c) and (d) of Figure 4 plot the estimated level and spread factors. The non-stationarity feature of the two factors is evident. Panel (b) of Figure 4 gives respectively the first three transitory factor loadings. With H_P and H_T estimated, we may decompose the time series of forward rate curves into its permanent and transitory components. This decomposition is presented in Panels (b) and (c) of Figure 3. This decomposition clearly separates the non-stationary component from the stationary component.

4.2 Identification of Shocks

To investigate the dynamics of the forward rate curves, we identify three structural shocks that drive the forward rate curves. The three structural shocks are defined in the subspace V spanned by the three leading functional principal components of the sample variance operator $\widehat{\Gamma}$. Denoting by Π the orthogonal projection on V, we let

$$\underline{\varepsilon}_t = \Pi \varepsilon_t$$

for $t = 1, \ldots, n$, and define the variance operator and sample variance operator of $(\underline{\varepsilon}_t)$ as $\underline{\Sigma} = \mathbb{E}(\underline{\varepsilon}_t \otimes \underline{\varepsilon}_t)$ and

$$\overline{\Sigma} = \frac{1}{n} \sum_{t=1}^{n} \left(\underline{\varepsilon}_t \otimes \underline{\varepsilon}_t \right),$$

respectively. For the fitted residuals $(\hat{\varepsilon}_t)$, the projected functional errors $(\hat{\underline{\varepsilon}}_t)$ with $\hat{\underline{\varepsilon}}_t = \widehat{\Pi}\hat{\varepsilon}_t$ for $t = 1, \ldots, n$ explain 93.3% of the total variation of $(\hat{\varepsilon}_t)$ over time, and therefore, most of the temporal fluctuations of the latter are captured by the former. Note that the variance operator and sample variance operator of $(\underline{\varepsilon}_t)$ are 3-dimensional whereas those of (ε_t) are infinite dimensional.

We identify two permanent shocks and one transitory shock: By definition, a permanent shock moves the forward rate curve everlastingly and a transitory shock shifts the forward rate curve only temporarily. The first shock will be identified to be a permanent shock that affects only the level of the forward rate curve in the long-run. The second shock will be identified to be a permanent shock that affects both the level and the spread of the curve in the long-run. In this spirit, we shall label these two shocks as the level shock and the spread shock, which are denoted as (e_t^L) and (e_t^S), respectively. The third one will be labeled as the transitory shock and denoted as (e_t^T).

To define our structural shocks more explicitly, we let $(v_k)_{k=1}^3$ span V such that v_1 and v_2 span H_N and v_3 is in H_S. Further, assuming a constant function is in H_N, we may let v_1 be a constant function and v_2 be orthogonal to the constant function without loss of generality. They are consistently estimable by the three leading functional principal components $(\hat{v}_k)_{k=1}^3$ of $\widehat{\Gamma}$. A consistent estimate for the first basis element v_1 may be obtained by projecting a constant function onto the space spanned by $\hat{v}_1$ and $\hat{v}_2$. The second basis element v_2 may then be consistently estimated by redefining $\hat{v}_2$ orthogonal to the estimated v_1 using the Gram-Schmidt procedure. Needless to say, v_3 is consistently estimated directly by $\hat{v}_3$.

Note that $\overline{\Sigma}$ is an operator on V spanned by $(v_k)_{k=1}^3$. Therefore, it may be written as $\overline{\Sigma} = \sum_{i,j=1}^3 \langle v_i, \overline{\Sigma} v_j \rangle (v_i \otimes v_j)$ and can be effectively represented as a 3-by-3 matrix $(\!|\overline{\Sigma}|\!)$ whose (i,j)-th element is given by $\langle v_i, \overline{\Sigma} v_j \rangle$ for $i,j = 1,2,3$. Similarly, $(\underline{\varepsilon}_t)$ takes values in V spanned by $(v_k)_{k=1}^3$, and therefore, it may be written as $\underline{\varepsilon}_t = \sum_{k=1}^3 \langle v_k, \underline{\varepsilon}_t \rangle v_k$ and can be effectively represented as a 3-dimensional vector $(\!|\underline{\varepsilon}_t|\!)$ whose k-th entry is given by $\langle v_k, \underline{\varepsilon}_t \rangle$. It follows that

$$(\!|\overline{\Sigma}|\!) = \frac{1}{n} \sum_{t=1}^n (\!|\underline{\varepsilon}_t|\!)(\!|\underline{\varepsilon}_t|\!)'.$$

Now we let

$$(\!|\overline{\Sigma}|\!) = LL',$$

and write

$$(\!|\underline{\varepsilon}_t|\!) = LQ \begin{pmatrix} e_t^L \\ e_t^S \\ e_t^T \end{pmatrix},$$

where L is a 3-by-3 lower triangular matrix, Q is a 3-by-3 orthogonal matrix, and (e_t^L), (e_t^S) and (e_t^T) are three structural shocks introduced above. Subsequently, we show that Q is uniquely defined and our structural errors are identified.

Let the (i,j)-th element of Q be (κ_{ij}) for $i,j = 1,2,3$. First, we find $\kappa_3 = (\kappa_{13}, \kappa_{23}, \kappa_{33})'$ satisfying $\|\kappa_3\| = 1$ and

$$L\kappa_3 = c \left(\!\left| \Pi_T v_3 \right|\!\right)$$

for some constant c, where $\Pi_T v_3$ is written as $\Pi_T v_3 = \sum_{k=1}^3 \langle v_k, \Pi_T v_3 \rangle v_k$ and represented by $\left(\!\left|\Pi_T v_3\right|\!\right)$ whose k-th entry is given by $\langle v_k, \Pi_T v_3 \rangle$. The shock (e_t^T) has at-impact response in H_T, and therefore, it only has a transitory effect on the forward rate curve. Second, for $\kappa_1 = (\kappa_{11}, \kappa_{21}, \kappa_{31})'$, we require $\|\kappa_1\| = 1$, $\kappa_1'\kappa_3 = 0$ and

$$L\kappa_1 = \begin{pmatrix} a \\ 0 \\ b \end{pmatrix}$$

for some constants a and b. Note that the shock (e_t^L) has at-impact response given by $av_1 + bv_3$. However, we have $A^h v_1 = v_1$ for any $h \geqslant 1$ and $A^h v_3 \to 0$ as $h \to \infty$, and therefore, (e_t^L) would have a long-run effect av_1 implying a shift in the level of the forward rate curve. Finally, we define $\kappa_2 = (\kappa_{12}, \kappa_{22}, \kappa_{32})'$ simply to satisfy $\|\kappa_2\| = 1$, $\kappa_2'\kappa_1 = 0$ and $\kappa_2'\kappa_3 = 0$, so that at-impact response of the shock (e_t^S) can be anywhere in V. It therefore has a long-run effect, which may change both the level and slope of the forward rate curve.

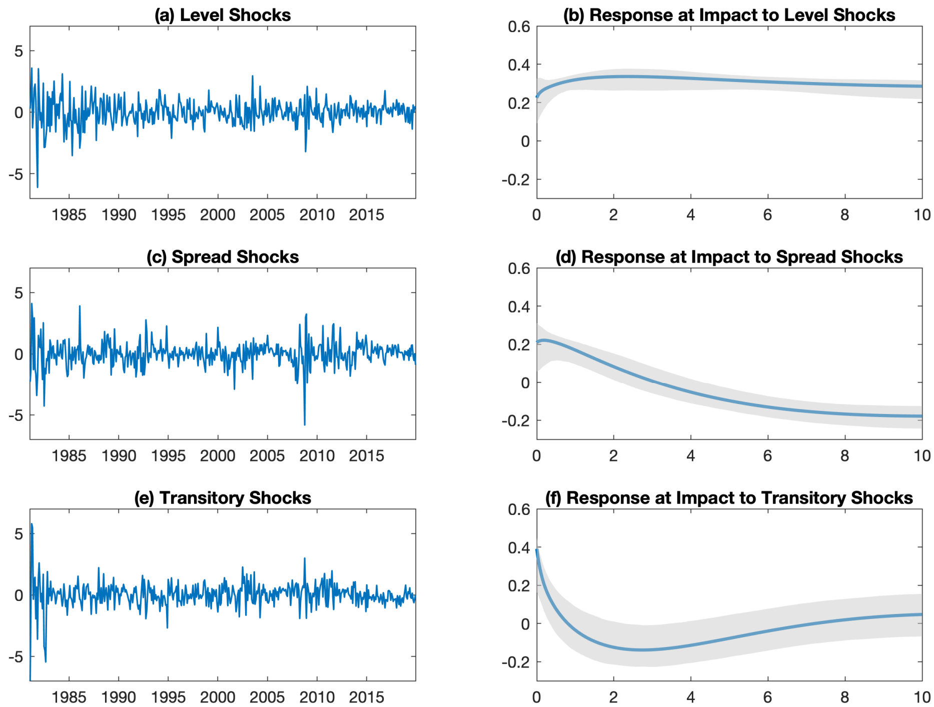
Once Q is identified, we let $P = LQ$, where P is a 3-by-3 matrix P with the (i,j)-th element given by (π_{ij}) for $i,j = 1,2,3$, and define the impulse response function

$$\mathrm{IRF}_i(h) = A^h \sum_{j=1}^{3} \pi_{ji} v_j = \sum_{j=1}^{3} \pi_{ji} A^h v_j,$$

where A is the autoregressive operator, h is the number of periods after the shock, and $i = 1,2,3$ corresponds to each of the three shocks $(e_t^L), (e_t^S)$ and (e_t^T), respectively. Since the signs of each column of P is not identified, we normalize the sign of the first column of P so that the response at impact to a positive level shock is positive, the second column of P so that the response at impact to a positive spread shock is downward sloping, and the third column of P so that the response at impact to a positive transitory shock has a trough at the maturity of two to three years. We normalize the signs of the structural shocks accordingly. For our impulse response analysis, we use the restricted version $\widetilde{A}$ in (17), instead of the unrestricted version $\widehat{A}$ in (15), as a consistent estimator for A.

The left three panels of Figure 5 plot the time series of the estimated three structural shocks, and the right three panels of Figure 5 give the response functions of the three structural shocks at impact. The level shock is a persistent shock that changes the forward rates uniformly at all maturities. The term premium shock is a persistent shock that affects the short forward rates and the long forward rates in opposite directions. A positive transitory shock increases the very short forward rates by a large amount, and the effect becomes negative very quickly as we increase the maturity, and it tends to die out when the maturity is very large.

Figure 5: Structural Forward Rate Shocks



Notes: The left three panels plot the time series of the three identified structural shocks, namely the level shocks, the spread shocks and the transitory shocks, in the forward rate curve dynamics, respectively. The right three panels plot the impulse response functions at impact to the level, spread, and transitory shocks respectively with their 95% bootstrap pointwise confidence bands based on 2000 repetitions.

4.3 Empirical Results

To investigate how monetary and fiscal policies affect the term structure dynamics, we look at the correlations between the structural forward rate shocks and the policy shocks. We also calculate the canonical correlations between the projected policy shocks and the three structural forward rate curve shocks altogether. The monetary policy shocks are obtained from Miranda-Agrippino and Ricco (2021) and the fiscal policy shocks are obtained from Romer and Romer (2010). We also consider other measures for policy shocks. See Appendix B.2 for details.

Table 1 presents the sample correlations and their significance levels obtained from the corresponding bootstrap distributions. We find that monetary policy shock has significant correlations with the permanent spread shock and the transitory shock. Fiscal policy shock,

Table 1: Correlations Between Policy Shocks and Structural Forward Rate Curve Shocks

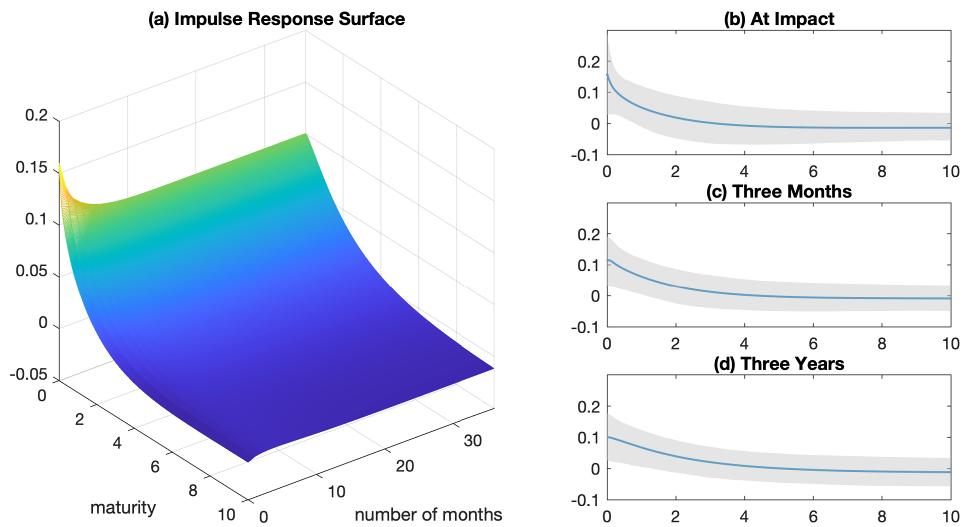
Correlations	Monetary	Fiscal
Level	0.036	-0.141^{*}
	[-0.070, 0.168]	[-0.293, 0.011]
Spread	0.251^{**}	-0.071
	[0.073, 0.402]	[-0.244, 0.203]
Transitory	$0.115^{\dagger}$	$-0.092^{\dagger}$
	[-0.063, 0.371]	[-0.326, 0.073]

Notes: This table presents the correlation coefficients between monetary and fiscal policy shocks and the three structural forward rate curve shocks. **,* and † denotes significance levels of 0.05, 0.1 and 0.32, respectively. The square brackets give the 95% bootstrap confidence intervals based on 2000 repetitions.

on the other hand, has significant correlations with the permanent level shock and the transitory shock. The canonical correlation between monetary policy shock and the three structural forward rate curve shocks is estimated to be 0.296 with a 95% confidence interval $[0.125, 0.504]$ obtained from its bootstrap distribution. The canonical correlation between fiscal policy shock and the structural forward rate curve shocks is estimated to be 0.185 with a 95% confidence interval $[0.093, 0.431]$ obtained from its bootstrap distribution.

We also look at the impulse responses of the forward rate curve to a monetary or fiscal policy shock. The idea is that policy shocks may induce structural forward rate curve shocks, at the scale of the corresponding correlations estimated above, which in turn drives changes in the forward rate curve dynamics over the horizons. The impulse responses are then given by linear combinations of the three impulse responses where the weights are given by the vector of the correlations between the policy shocks and the structural forward rate curve shocks. Panel (a) of Figure 6 presents the estimated impulse response surface of the forward rate curve to a monetary policy shock up to 36 months after the initial shock. Panels (b), (c) and (d) of Figure 6 plot the impulse responses at impact, three months, and three years after the shock, respectively, with the pointwise 95% bootstrap confidence bands. The bootstrapped impulse responses of policy shocks are calculated from the bootstrapped impulse responses of the structural forward rate curve shocks and the bootstrapped correlations. Similarly, Panel (a) of Figure 7 presents the estimated impulse response surface of the forward rate curve to a fiscal policy shock up to 36 months after the initial shock. Panels (b), (c) and (d) of Figure 7 plot the impulse responses at impact, three months, and three years after the shock, respectively, with the pointwise 95% bootstrap confidence bands. The effect of monetary policy shocks to the forward rate curve dynamics is significant at very short terms. The overall long term response of the forward rate curve to a positive monetary policy shock (a rise in unexpected federal funds rate) is estimated to

Figure 6: Impulse Responses of Forward Rate Curves to Monetary Policy Shocks with Bands



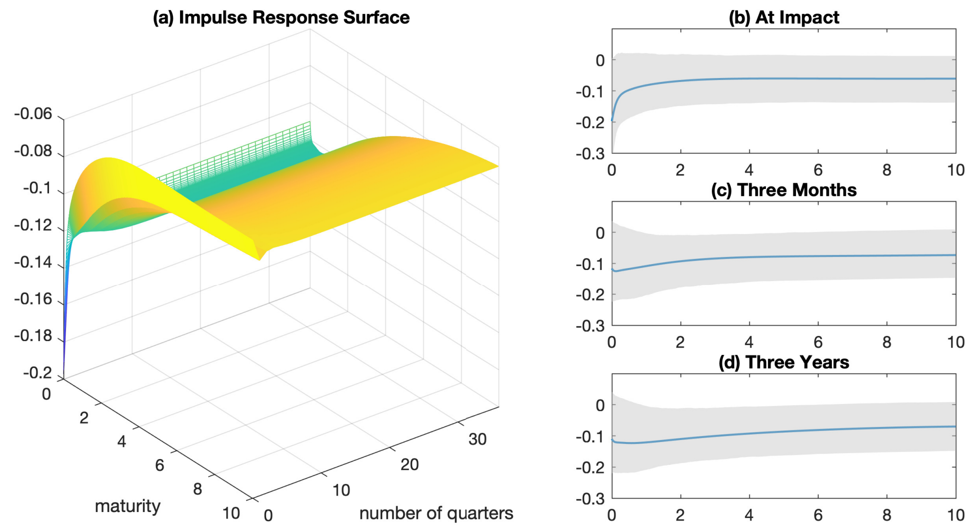
Notes: The left panel presents the functional impulse response surface of the forward rate curve to a positive monetary policy shock up to 36 months after the initial impact. The right three panels present the functional impulse responses at impact, three months, and three years after a monetary policy shock, respectively, with the pointwise 95% bootstrap confidence bands.

be positive, but not statistically significant at the level of 0.05. The overall long term effect of a positive fiscal policy shock (a rise in tax) on the forward rate curve is estimated to be negative, and is statistically significant at the significance level of 0.05. By a similar exercise, we may analyze any other feature of policy consequences regarding the term structure of interest rates both in the short run and in the long run by investigating the functional impulse responses of the forward rate curve to monetary and fiscal policy shocks.

5 Conclusions

We build an autoregressive model for time series of random functions taking values in a Hilbert space with persistence. A process generated by this model admits a decomposition into a permanent component and a transitory component, representing the persistent stochastic trend and the stationary cyclical movement in the process, respectively. We show how to estimate the model, both without and with the unit root restriction, and how to conduct decompositions and make predictions. The estimated autoregressive operator is consistent under very mild conditions with different convergence rates and limit distribu-

Figure 7: Impulse Responses of Forward Rate Curves to Fiscal Policy Shocks with Bands



Notes: The left panel presents the functional impulse response surface of the forward rate curve to a positive fiscal policy shock up to 36 months after the initial impact. The right three panels present the functional impulse responses at impact, three months, and three years after a positive fiscal policy shock, respectively, with the pointwise 95% bootstrap confidence bands.

tions in different subspaces, and the predictor is asymptotically normal, with a convergence rate slower than the usual $\sqrt{n}$ rate. We extend our baseline model to the case in which the transitory component has a non-zero drift term, the time series of functions is estimated with error, and the functional process follows a general autoregressive process. We apply our method to study the term structure of the US government bond yields. We decompose the forward rate curve series into its permanent and transitory components, identify two permanent structural shocks and one transitory structural shocks that drive the forward rate curve dynamics, and find that monetary and fiscal policies are correlated with these structural shocks. We give the impulse response surfaces of policy shocks to forward rate curve dynamics.

Appendix A Extensions

A.1 Model with Nonzero Drift

It is rather straightforward to extend our framework to allow for the existence of transitory component with nonzero mean in functional autoregression with unit roots. To show how, we consider the functional autoregression

$$f_t = \nu + Af_{t-1} + \varepsilon_t \tag{19}$$

in place of (1), where $\nu \in H_T$ and A satisfies Assumptions 2.1 and 2.2. Note that ν is assumed to be in H_T, and therefore, it introduces a drift term only in the transitory component (f_t^T) of (f_t). As is well known, the presence of a non-zero drift term in the permanent component (f_t^P) of (f_t) would generate a linear time trend in (f_t). We may rewrite the functional autoregressive model in (19) as

$$f_t - \mu = A(f_{t-1} - \mu) + \varepsilon_t$$

where $\mu = \mathbb{E}f_t^T = (1 - A_T)^{-1}\nu$ is in H_T.

To estimate the autoregressive operator A in (19), we need to first demean the time series (f_t), where the sample mean of the time series is given by

$$\bar{f} = \frac{1}{n}\sum_{t=1}^{n} f_t.$$

Subsequently, we denote the demeaned time series of (f_t) by $f_t^\mu = f_t - \bar{f}$, and redefine the operator $\widehat{\Gamma}$ by

$$\widehat{\Gamma} = \sum_{t=1}^{n}(f_t^\mu \otimes f_t^\mu),$$

and redefine $\hat{\lambda}_i$ and $\hat{v}_i$ as the ordered eigenvalues and eigenfunctions of the newly defined $\widehat{\Gamma}$.

Due to Corollary 3.2 in Bosq (2000), we have

$$\left\| \frac{1}{n}\sum_{t=1}^{n} f_t^T - \mathbb{E}f_t^T \right\| = O(n^{-1/2}\log^{1/2} n) \quad \text{a.s.},$$

for large n, and consequently, on H_T, all our sample statistics redefined by (f_t^μ) yield the same asymptotics as those defined for (f_t). Therefore, the use of (f_t^μ) gets rid of the nonzero mean in (f_t) without affecting any asymptotics in H_T. On the contrary, however, demeaning (f_t^μ) does affect asymptotics in H_P. More precisely, our previous asymptotics involving functions of W_P are now replaced by functions of

$$W_P^\mu(r) = W_P(r) - \int_0^1 W_P(r)dr,$$

i.e., the demeaned Brownian motion on H_P. The interested reader is referred to Section 4 of Chang et al. (2016b) for more details.

It is straightforward to establish the following lemma, which is analogous to Lemma 3.2.

Lemma A.1. *Let Assumptions 2.1 and 2.2 hold. Then*

$$\overline{\Gamma}_{NN} \to_d \int_0^1 (W_P^\mu \otimes W_P^\mu)(r)dr$$
$$\overline{\Gamma}_{SS} \to_p \Pi_S \left(\sum_{k=0}^\infty A_T^k \Sigma_T A_T^{*k} \right) \Pi_S$$

as $n \to \infty$. Moreover, we have

$$\overline{\Gamma}_{NS} = \overline{\Gamma}_{SN}^* = O_p(1)$$

for large n.

Lemma C.2 continues to hold for the functional autoregression (19). Moreover, Theorem 3.3 holds with Γ_{NN} redefined as $\Gamma_{NN} = \int_0^1 (W_P^\mu \otimes W_P^\mu)(r)dr$, and Lemma C.1 holds with (f_t) replaced by the (f_t^μ). The following theorem shows that the demeaning procedure does not affect the asymptotic properties of our FAR estimator and predictor. The predictor of course should be modified as

$$\hat{f}_{n+1} = \bar{f} + \widehat{A} f_n^\mu$$

to reflect the required demeaning procedure.

Theorem A.2. *Let the assumptions in Theorem 3.4 hold. Then*

$$\left\| \widehat{A} - A \right\| \to_p 0$$

and

$$\left\| \widetilde{A} - A \right\| \to_p 0$$

as $n \to \infty$. If in addition the assumptions in Theorem 3.11 hold, then

$$\sqrt{n/m} \left((\hat{f}_{n+1} - \mu) - A(f_n - \mu) \right) \to_d \mathbb{N}(0, \Sigma)$$

as $n \to \infty$.

In sum, when the stationary component is not mean-zero and a demeaning procedure is required, the asymptotic results developed for the estimation and prediction of mean-zero functional autoregression continue to hold essentially without any additional assumptions.

A.2 Regression with Estimated Functional Time Series

In virtually all practical applications, we expect that (f_t) is not directly observable and has to be estimated from either cross-sectional or high-frequency observations. In this case, we may analyze our functional autoregressive model using the estimated functional time series $(\hat{f}_t)$. It is also possible to allow for the presence of drift term in the stationary component of (f_t), in which case we may use $(\hat{f}_t^{\mu})$,

$$\hat{f}_t^{\mu} = \hat{f}_t - \frac{1}{n}\sum_{t=1}^{n}\hat{f}_t$$

in place of $(\hat{f}_t)$.

We denote the estimation error of f_t by $\Delta_t = \hat{f}_t - f_t$. In order to preserve our asymptotic results as in Section 3, we need to control the magnitude of (Δ_t). We therefore introduce the following assumption.

Assumption A.1. $\sup_{t\geqslant 1}\|\Delta_t\| = O_p(1/\sqrt{n})$.

Under Assumption A.1, $\|\Delta_t\|$ becomes negligible uniformly in $t = 1, 2, \ldots$, and all our asymptotic results based on (f_t) continue to hold also for $(\hat{f}_t)$. The use of estimated functions, in place of the true functions, therefore has no bearing on our asymptotics. This is well expected from Chang et al. (2016b). The condition required here is not absolutely necessary and can be relaxed if we introduce some additional assumptions. However, it is already not stringent and expected to hold as long as the number of observations we use to obtain $(\hat{f}_t)$ is sufficiently large compared with n, which appears to be the case for many practical applications.

A.3 Higher Order Autoregression

In this section, we consider the functional autoregression model of order $p > 1$ with unit roots. Suppose that (f_t) follows an FAR(p) model given by

$$f_t = A_1 f_{t-1} + A_2 f_{t-2} + \cdots + A_p f_{t-p} + \varepsilon_t, \tag{20}$$

where $A_1, A_2, \cdots, A_p$ are compact operators on H and (ε_t) is a functional white noise that satisfies (c) of Assumption 2.1.

Consider the direct sum $H^p = H \oplus \cdots \oplus H$ equipped with the inner product defined by $\langle (u_1, \cdots, u_p), (v_1, \cdots, v_p) \rangle = \sum_{i=1}^p \langle u_i, v_i \rangle$ for all $v_i \in H$ and $u_i \in H$. We may rewrite the FAR(p) process in (20) as an H^p-valued FAR(1) process given by

$$g_t = Bg_{t-1} + \eta_t \tag{21}$$

where $g_t = (f_t, f_{t-1}, \cdots, f_{t-p+1}), \eta_t = (\varepsilon_t, 0, \cdots, 0)$ and

$$B = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

We define the characteristic polynomial $A(z) = z^p - z^{p-1}A_1 - \cdots - zA_{p-1} - A_p$ for $z \in \mathbb{C}$ and introduce the following assumption.

Assumption A.2. $A(1)$ is not invertible, and if $A(z)$ is not invertible, then $z = 1$ or $|z| < 1$.

Define

$$M(z) = \begin{bmatrix} 0 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -1 \\ A_0(z) & A_1(z) & A_2(z) & \cdots & A_{p-2}(z) & A_{p-1}(z) \end{bmatrix}$$

where $A_0(z) = 1$ and $A_i(z) = zA_{i-1}(z) - A_i$ for $i \geqslant 0$, and define

$$N(z) = \begin{bmatrix} 1 & z & z^2 & \cdots & z^{p-2} & z^{p-1} \\ 0 & 1 & z & \cdots & z^{p-3} & z^{p-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & z \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

Then we have that

$$M(z)(zI^p - B)N(z) = \begin{bmatrix} 1 & 0 \\ 0 & A(z) \end{bmatrix}.$$

By construction, $M(z)$ and $N(z)$ are invertible for all $z \in \mathbb{C}$, and therefore we have

$$\lambda(A) = \{z : A(z) \text{ is not invertible}\}.$$

However, since $\lambda(A)$ is closed and 1 cannot be a limit point of $\lambda(A)$, Assumption A.2 implies that $\sup \lambda(A)\backslash\{1\} < 1$. Furthermore, since $\sup \lambda(A)\backslash\{1\} = \lim_{r\to\infty} \|A_T^r\|^{1/r}$, there exists $r \in \mathbb{N}$ such that $\|A_T^r\| < 1$. Consequently, Assumption 2.2 holds for the model (21). This suggests that whenever we have an FAR(p) model with unit roots, we may reformulate it as an FAR(1) process and therefore all theoretical results for the FAR(1) model remain valid for the FAR(p) model.

To estimate the FAR(p) model, we may write it in the form of (21) and conduct estimation based on FAR(1), or we may estimate a finite dimensional version of the FAR(p) model given by

$$(f_t) = (A_1)(f_{t-1}) + \cdots + (A_p)(f_{t-p}) + (\varepsilon) \tag{22}$$

where (f_t) is the m-dimensional vector whose elements are $\langle f_t, v_k\rangle$ for $k = 1, \ldots, m$, and $(A_1), \ldots, (A_p)$ are $m \times m$ coefficient matrices. In actual estimation, we replace v_k by $\hat{v}_k$. Once the estimate $\widehat{(A_k)}$ of (A_k) is obtained, we recover the estimate $\widehat{A_k}$ of A_k by $\widehat{A_k} = \sum_{i,j=1}^m \widehat{(A_k)}_{ij}(v_i \otimes v_j)$ where $\widehat{(A_k)}_{ij}$ is the (i,j)-th entry of $\widehat{(A_k)}$.

To implement the unit root restriction in the FAR(p) setting, we could first run two auxiliary regressions

$$\Delta f_t = \Theta_1 \Delta f_{t-1} + \Theta_2 \Delta f_{t-2} + \cdots + \Theta_{p-1} \Delta f_{t-p+1} + u_t$$

and

$$f_{t-1} = \Xi_1 \Delta f_{t-1} + \Xi_2 \Delta f_{t-2} + \cdots + \Xi_{p-1} \Delta f_{t-p+1} + w_t$$

and obtain the estimates $\widehat{\Theta}_k, \widehat{\Xi}_k$ and residuals $\hat{u}_t$ and $\hat{w}_t$. The two regressions can be conducted based on their finite dimensional versions as in (22). We then construct the estimators

$$\widehat{\Sigma}_{uu} = \frac{1}{n} \sum_{t=1}^n \hat{u}_t \otimes \hat{u}_t,$$

$$\widehat{\Sigma}_{ww} = \frac{1}{n} \sum_{t=1}^n \hat{w}_t \otimes \hat{w}_t,$$

$$\widehat{\Sigma}_{uw} = \frac{1}{n} \sum_{t=1}^n \hat{u}_t \otimes \hat{w}_t,$$

and

$$\widehat{\Sigma}_{wu} = \frac{1}{n} \sum_{t=1}^{n} \hat{w}_t \otimes \hat{u}_t.$$

We get the $m - \ell$ largest eigenvalues $\hat{\kappa}_1, \ldots, \hat{\kappa}_{m-\ell}$ and the corresponding eigenvectors $\hat{\phi}_1, \ldots, \hat{\phi}_{m-\ell}$ of the operator $\widehat{\Sigma}_{ww}^{+} \widehat{\Sigma}_{wu} \widehat{\Sigma}_{uu}^{+} \widehat{\Sigma}_{uw}$, where $^{+}$ denotes the pseudo-inverse in the span of $(\hat{v}_k), k = 1, \ldots, m$. We construct

$$\widehat{\Psi}_0 = \widehat{\Sigma}_{uw} \left(\sum_{k=1}^{m-\ell} \hat{\phi}_k \otimes \hat{\phi}_k \right)$$

and

$$\widehat{\Psi}_k = \widehat{\Theta}_k - \widehat{\Psi}_0 \widehat{\Xi}_k,$$

for $k = 1, \ldots, p-1$. Then $A_1, \ldots, A_p$ can be estimated by $\widehat{A}_1 = \widehat{\Pi} + \widehat{\Psi}_0 + \widehat{\Psi}_1$, $\widehat{A}_k = \widehat{\Psi}_k - \widehat{\Psi}_{k-1}$ for $k = 2, \ldots, p-1$, and $\widehat{A}_p = -\widehat{\Psi}_{p-1}$.

Appendix B Supplements to the Empirical Analysis of Term Structure

B.1 Implementation and the Truncation Parameter

To obtain the estimator, after we get the forward rate curves, we demean the curves by subtracting the sample mean from the time series of forward rate curves and then represent the demeaned forward rate curve in each period with the Daubechies wavelets using 1037 basis functions. That is, each forward rate curve is represented as a 1037-dimensional vector whose coordinates are the wavelet coefficients of the curve. We obtain the matrix representation of the operator $\widehat{\Gamma}$ as a 1037-by-1037 matrix, and obtain the eigenvalues and eigenvectors of $\widehat{\Gamma}$. Note that the eigenvectors can be transformed to eigenfunctions using the wavelet basis. With m given, we then estimate A.

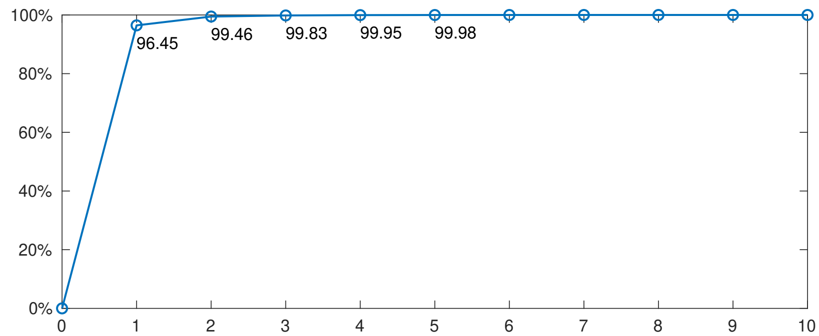
To settle down the value of m, we split the sample, using the first 4/5 of the sample to estimate the model and the last 1/5 of the sample to conduct out-of-sample prediction and select the value of m that yields the best prediction performance. It turns out that the best value of m is 5. This implies that we are going to include the first five principal components in our analysis.

Panel (d) of Figure 3 gives the component of the forward rate curve process that is not included in the first five principal components. This component is negligible, indicating that our approximation is precise. Figure 8 plots the cumulative ratios of the ten largest eigenvalues of the unnormalized sample variance operator $\widehat{\Gamma}$ to the sum of all eigenvalues of $\widehat{\Gamma}$. It is well known from the theory of principal component analysis that the ratio of an eigenvalue to the sum of eigenvalues gives the proportion of data variance that is explained by the corresponding principal component. Figure 8 shows that the first five principal components explain 99.98% of variations in the data, which justifies our choice of the value of m.

B.2 Alternative Measures of Policy Shocks

Besides the Miranda-Agrippino and Ricco (2021) monetary policy shocks, we also consider the Romer and Romer (2004) monetary policy shocks as an alternative. The Romer and Romer (2004) monetary policy shocks are constructed as the residuals from projecting the Federal Reserves' intended changes in the federal funds rate on the forecast of economic growth, inflation, and unemployment. Their data are available monthly from January 1966 to December 1996. We use their data from January 1981 to December 1996 to align with our forward rate curve data.

Figure 8: Cumulative Scree Plot of the Forward Rate Curves



Notes: This figure plots the cumulative proportions of data variance that are explained by the first ten principal components. These proportions are calculated as the cumulative ratios of the ten largest eigenvalues of the unnormalized sample variance operator $\widehat{\Gamma}$ to the sum of all eigenvalues of $\widehat{\Gamma}$.

The Romer and Romer (2010) fiscal policy shocks used in our empirical application are constructed from the narrative records of tax policy actions. For each tax policy change record, the authors determine the motivation, timing, and size of the tax change, and use those time and size as the time and value of the shock. For our analysis, we use the combined tax changes they provide, which include both the endogenous tax changes used to boost growth in the near future and the exogenous changes used for other purposes. The fiscal policy shock data are available from the first quarter of 1945 to the last quarter of 2007. Since the fiscal policy shocks are available quarterly, we apply our first order functional model to quarterly forward rate curves from the first quarter of 1981 to the last quarter of 2007.

Besides using the original policy shocks directly, we propose the following model $z_t^o = z_t + \varepsilon_t^z = \beta'\varepsilon_t^\eta + \varepsilon_t^z$ where z_t^o is the original policy shocks (either monetary or fiscal), z_t is the projected policy shocks defined as the projection of the unadjusted policy shocks onto $\varepsilon_t^\eta = (\varepsilon_t^{\eta f}, \varepsilon_t^{\eta x})'$, which consists of a collection of innovations $\varepsilon_t^{\eta f}$ to the common factors and innovations $\varepsilon_t^{\eta x}$ to idiosyncratic components of a large set of macroeconomic variables, and ε_t^z is the error term associated with the policy shocks. To obtain the innovations ε_t^η, we consider the following factor model $x_{it} = \sum_{j=1}^{J} \lambda_{ij}\eta_{jt}^f + \eta_{it}^x$ where x_{it} is the i-th standardized macroeconomic variable at time t, $t = 1, 2, \ldots T, i = 1, 2, \ldots, N$, η_{jt}^f is the j-th common factor with loading λ_{ij}, $j = 1, 2, \ldots, J$, and η_{it}^x is the idiosyncratic component specific to the i-th macroeconomic variable for $i = 1, 2, \ldots, N$. We assume that the J-variate common factor $\eta_t^f = (\eta_{1t}^f, \ldots, \eta_{Jt}^f)'$ jointly follow a J-variate VAR process and each of the N individual idiosyncratic component $\{\eta_{it}^x\}$ follow an AR(p_i) process. We call the innovations $\varepsilon_t^{\eta f}$ to the VAR process $\{\eta_t^f\}$ the common factor innovations, and the innovations $\varepsilon_t^{\eta x} =$

Table 2: Correlations Between Policy Shocks and Structural Forward Rate Curve Shocks

Correlations	Monetary Policy Shock		Fiscal Policy Shock	
	Projected	Original	Projected	Original
Level	0.023	−0.010	−0.141*	−0.141*
	[-0.133, 0.168]	[-0.137, 0.164]	[-0.293, 0.011]	[-0.293, 0.011]
Spread	0.299*	0.289*	−0.071	−0.071
	[-0.139, 0.490]	[-0.135, 0.487]	[-0.244, 0.203]	[-0.244, 0.203]
Transitory	0.091	0.074	$-0.092^{\dagger}$	$-0.092^{\dagger}$
	[-0.132, 0.358]	[-0.133, 0.353]	[-0.326, 0.073]	[-0.326, 0.073]

Notes: This table presents the correlation coefficients between Romer and Romer (2004) monetary policy shocks and Romer and Romer (2010) fiscal policy shocks and the three structural forward rate curve shocks. The results obtained using both the projected policy shocks and the original policy shocks are reported. **,* and † denotes significance levels of 0.05, 0.1 and 0.32, respectively. The square brackets give the 95% bootstrap confidence intervals based on 2000 repetitions.

$(\varepsilon_{1t}^{\eta x}, \ldots, \varepsilon_{Nt}^{\eta x})'$ to the individual AR(p_i) processes $\eta_t^x = (\eta_{1t}^x, \ldots, \eta_{Nt}^x)$ the macroeconomic innovations. For $\{x_{it}\}$, we use the set of variables in the FRED-MD/FRED-QD database developed by McCracken and Ng (2016), which contains 127 macroeconomic variables at monthly frequency and 236 macroeconomic variables at quarterly frequency. We include three factors chosen by the eigenvalue ratio test, and the orders of the VAR and AR models are determined by BIC.

There are at least two merits of using projected policy shocks rather than the original policy shocks. First, projecting the policy shocks onto the span of the macroeconomic and common factor innovations purges the indirect effects of policy shocks to the forward rate curves that work through first affecting macroeconomic variables. Second, the projection provides a way for us to interpolate and extrapolate in case we have missing values in the original policy shocks data, although in our analysis we do not deal with data missing values.

Table 2 presents the sample correlations between the Romer and Romer (2004) monetary policy shocks and the Romer and Romer (2010) fiscal policy shocks and the three structural forward rate curve shocks. For the shocks, both the original version and the projected version are used. It turns out that using the projected shocks instead of the original shocks does not change the results in any statistically significant way. The results using these alternative policy shocks turn out to be similar comparing to the results in Table 1 in the main text. Fiscal policy shock has significant correlations with the permanent level shock and the transitory shock. Monetary policy shock has significant correlations with the permanent spread shock, although not with the transitory shocks.

Appendix C Mathematical Proofs

The following lemma, providing the orders of the interaction terms, is useful in the proof of asymptotics $\widehat{A}$ defined in (15).

Lemma C.1. *Let Assumptions 2.1 and 2.2 hold. Then*

$$\left\| \sum_{t=1}^{n} (\varepsilon_t \otimes f_{t-1}^S) \right\| = O_p(\sqrt{n}) \quad \textit{and} \quad \left\| \sum_{t=1}^{n} (\varepsilon_t \otimes f_{t-1}^N) \right\| = O_p(n)$$

for large n.

The terms in the above theorem have orders that are the same as their finite dimensional counterparts. See, for example, Lemma 2.1 in Park and Phillips (1988).

Also, the following lemma from Hu et al. (2016) is useful.

Lemma C.2. *Let Assumptions 2.1 and 2.2 hold. Then*

$$\left\| \overline{\Gamma}_{SS} - \Gamma_{SS} \right\| = O\left(n^{-1/2} \log^{1/2} n \right) \quad a.s..$$

Moreover,

$$\sup_{k \geqslant \ell+1} \left| n^{-1} \hat{\lambda}_k - \lambda_k \right| \leqslant \left\| \overline{\Gamma}_{SS} - \Gamma_{SS} \right\|$$

and

$$\| \hat{v}_k - v_k \| \leqslant \tau_k \left\| \overline{\Gamma}_{SS} - \Gamma_{SS} \right\|$$

for $k = \ell+1, \ell+2, \ldots$.

Proof of Theorem 2.1. We use tools from functional calculus in this proof. We refer interested readers to Gohberg et al. (1990), in particular section I.1, I.2, II.1 and II.3, for details.

Since A is a compact operator on a separable Hilbert space H, $\lambda(A)$ is at most countable and could have only 0 as a limit point. This implies that we may separate $\{1\}$ from the other elements in $\lambda(A)$ by two non-intersecting Cauchy contours Γ_P and Γ_T specified in Section 2.2. It follows from Lemma 2.1, Theorem 2.2 and Corollary 2.3 in Chapter 1 of Gohberg et al. (1990) that $\Pi_P + \Pi_T = 1$, $\Pi_P \Pi_T = \Pi_T \Pi_P = 0$, Π_P is the projection onto the subspace H_P with kernel H_T, Π_T is the projection onto the subspace H_T with kernel H_P, $H = H_P \oplus H_T$, and that the two subspaces H_P and H_T are invariant with respect to A. Also, since all non-zero elements in $\lambda(A)$ are eigenvalues of finite type of A, we have that H_P is finite dimensional. $\square$

Proof of Lemma 2.2. Clearly, $1 - A_P$ is nilpotent of degree d on H_P if and only if $1 - A_P^*$ is nilpotent of degree d on H_P^*. However, $1 - A_P^*$ is nilpotent of degree d on H_P^* if and only if there is a basis including $v, (A_P^* - 1)v, \ldots, (A_P^* - 1)^{d-1}v$ for H_P^*, with some $v \in H_P^*$ such that $v \neq 0$, as shown in Theorems 1 and 2 of Section 57 in Halmos (1974).

Since $A_P^* - 1$ is nilpotent of degree d, we have

$$(A_P^* - 1)^d = A_P^*(A_P^* - 1)^{d-1} - (A_P^* - 1)^{d-1} = 0,$$

and therefore,

$$\begin{aligned}\Big\langle (A_P^* - 1)^{d-1}v, f_t^P \Big\rangle &= \Big\langle (A_P^* - 1)^{d-1}v, A_P f_{t-1}^P \Big\rangle + \Big\langle (A_P^* - 1)^{d-1}v, \varepsilon_t^P \Big\rangle \\ &= \Big\langle (A_P^* - 1)^{d-1}v, f_{t-1}^P \Big\rangle + \Big\langle (A_P^* - 1)^{d-1}v, \varepsilon_t^P \Big\rangle,\end{aligned}$$

which implies that $\big(\langle (A_P^* - 1)^{d-1}v, f_t^P \rangle\big)$ is I(1).

For $d \geqslant 2$, however, we have

$$A_P^*(A_P^* - 1)^{d-2} = (A_P^* - 1)^{d-2} + (A_P^* - 1)^{d-1},$$

from which it follows that

$$\begin{aligned}\Big\langle (A_P^* - 1)^{d-2}v, f_t^P \Big\rangle &= \Big\langle (A_P^* - 1)^{d-2}v, A_P f_{t-1}^P \Big\rangle + \Big\langle (A_P^* - 1)^{d-2}v, \varepsilon_t^P \Big\rangle \\ &= \Big\langle (A_P^* - 1)^{d-2}v, f_{t-1}^P \Big\rangle + \Big\langle (A_P^* - 1)^{d-1}v, f_{t-1}^P \Big\rangle \\ &\quad + \Big\langle (A_P^* - 1)^{d-2}v, \varepsilon_t^P \Big\rangle.\end{aligned}$$

This shows that $\big(\langle (A_P^* - 1)^{d-2}v, f_t^P \rangle\big)$ is I(2). By the usual mathematical induction, we may now readily show that $\big(\langle v, f_t^P \rangle\big)$ is I(d), and the proof is complete. □

Proof of Lemma 3.1. It follows from Theorem 2.7 in Bosq (2000) that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t \to_d \mathbb{N}(0, \Sigma)$$

where $\mathbb{N}(0, \Sigma)$ is an H-valued Gaussian random element with variance operator Σ. The invariance principle then follows immediately from Corollary 1 in Kuelbs (1973).

□

Proof of Lemma 3.2. See Lemma 3.1 in Chang et al. (2016b). □

Proof of Theorem 3.3. See Theorem 3.3 in Chang et al. (2016b). $\square$

Proof of Lemma C.1. Let B be the closed unit ball in H.

$$\begin{aligned}\left\|\sum_{t=1}^{n}(\varepsilon_t \otimes f_{t-1}^{S})\right\| &= \sup_{v_1,v_2\in B}\left|\left\langle v_1, \left(\sum_{t=1}^{n}(\varepsilon_t \otimes f_{t-1}^{S})\right) v_2\right\rangle\right| \\ &= \sup_{v_1,v_2\in B}\left|\sum_{t=1}^{n}\langle v_1,\varepsilon_t\rangle\langle v_2, f_{t-1}^{S}\rangle\right|.\end{aligned}$$

Note that for any v_1 and v_2 in H, $(\langle v_1,\varepsilon_t\rangle\langle v_2, f_{t-1}^{S}\rangle)$ is a martingale difference sequence, then by the central limit theorem for martingale difference sequence,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\langle v_1,\varepsilon_t\rangle\langle v_2, f_{t-1}^{S}\rangle \to_d \mathbb{N}\left(0, V_S(v_1,v_2)\right)$$

where

$$\begin{aligned}V_S(v_1,v_2) &= \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\langle v_1,\varepsilon_t\rangle^2\langle v_2, f_{t-1}^{S}\rangle^2 \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\left(\mathbb{E}\langle v_1,\varepsilon_t\rangle^2\right)\left(\mathbb{E}\langle v_2, f_{t-1}^{S}\rangle^2\right) \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\langle v_1, \mathbb{E}(\varepsilon_t\otimes\varepsilon_t)v_1\rangle\langle v_2, \mathbb{E}(f_{t-1}^{S}\otimes f_{t-1}^{S})v_2\rangle \\ &\leqslant \|\mathbb{E}(\varepsilon_t\otimes\varepsilon_t)\|\,\|\mathbb{E}(f_{t-1}^{S}\otimes f_{t-1}^{S})\|\end{aligned}$$

for all $v_1, v_2 \in B$. Since (ε_t) is a functional white noise and (f_t^T) is stationary, $V_S(v_1,v_2)$ is uniformly bounded (for $v_1, v_2 \in B$) by a constant. Therefore, the family of random operators $(1/\sqrt{n}\sum_{t=1}^{n}(\varepsilon_t\otimes f_{t-1}^{S}))$ is stochastically pointwise bounded. By a random Banach-Steinhaus theorem due to Velasco and Villena (1995), stochastic pointwise boundedness implies stochastic equicontinuity. Therefore we have that

$$\left\|\sum_{t=1}^{n}(\varepsilon_t \otimes f_{t-1}^{S})\right\| = O_p(\sqrt{n}).$$

Similarly, we have that

$$\left\|\sum_{t=1}^{n}(\varepsilon_t \otimes f_{t-1}^{N})\right\| = \sup_{v_1,v_2\in B}\left|\sum_{t=1}^{n}\langle v_1,\varepsilon_t\rangle\langle v_2, f_{t-1}^{N}\rangle\right|.$$

By Lemma 3.1 and the remarks that follows, we have

$$\frac{1}{n}\sum_{t=1}^{n}\langle v_1, \varepsilon_t\rangle\langle v_2, f_{t-1}^N\rangle \to_d \int_0^1 \langle v_2, W(r)\rangle d\langle v_1, W_P(r)\rangle.$$

The limiting distribution is a normal mixture, which is stochastically bounded. Once again this stochastic pointwise boundedness implies stochastic equicontinuity. That is,

$$\left\|\sum_{t=1}^{n}(\varepsilon_t \otimes f_{t-1}^N)\right\| = O_p(n).$$

□

Proof of Theorem C.2. See Theorem 2, Lemma 3 and Theorem 4 in Hu et al. (2016). □

Proof of Theorem 3.4. We first prove consistency. Write

$$\widehat{A} - A = \left(\widehat{A}\widehat{\Pi}_N - A\widehat{\Pi}_N\right) + \left(\widehat{A}\widetilde{\Pi}_S - A\widetilde{\Pi}_S\right).$$

First, note that

$$\widehat{A}\widehat{\Pi}_N - A\widehat{\Pi}_N = \left(\sum_{t=1}^{n}(\varepsilon_t \otimes f_{t-1})\right)\left(\sum_{k=1}^{\ell}\hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k)\right).$$

Since

$$\left\|\sum_{t=1}^{n}(\varepsilon_t \otimes f_{t-1})\right\| \leqslant \left\|\sum_{t=1}^{n}(\varepsilon_t \otimes f_{t-1}^S)\right\| + \left\|\sum_{t=1}^{n}(\varepsilon_t \otimes f_{t-1}^N)\right\| = O_p(n), \tag{23}$$

and that

$$\hat{\lambda}_k^{-1} = O_p(n^{-2}) \tag{24}$$

for all $k = 1, \ldots, \ell$, we have that

$$\left\|\widehat{A}\widehat{\Pi}_N - A\widehat{\Pi}_N\right\| = O_p(n^{-1}). \tag{25}$$

Next, we first show that

$$\left\|\sum_{k=\ell+1}^{m} n\hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) - \sum_{k=\ell+1}^{m} \lambda_k^{-1}(v_k \otimes v_k)\right\| = o_p(1). \tag{26}$$

Write

$$
\begin{aligned}
&\sum_{k=\ell+1}^{m} n\hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) - \sum_{k=\ell+1}^{m} \lambda_k^{-1}(v_k \otimes v_k) \\
&= \sum_{k=\ell+1}^{m} \left(n\hat{\lambda}_k^{-1} - \lambda_k^{-1}\right)(\hat{v}_k \otimes \hat{v}_k) + \sum_{k=\ell+1}^{m} \lambda_k^{-1}\big((\hat{v}_k \otimes \hat{v}_k) - (v_k \otimes v_k)\big).
\end{aligned}
$$

Since$\sum_{k=\ell+1}^{m} \tau_k \geqslant 2\sqrt{2}(\lambda_m - \lambda_{m+1})^{-1} \geqslant 2\sqrt{2}\lambda_m^{-1}$, by assumption we have that $\frac{\log n\left(\sum_{k=\ell+1}^{m} \tau_k\right)^2}{n\lambda_m^2} \geqslant \frac{8\log n}{n\lambda_m^4} \to 0$ as $n \to \infty$. This implies that

$$
\lambda_m^{-1} = o\left(n^{1/4}\log^{-1/4} n\right). \tag{27}
$$

For large enough k, we have that $n^{-1}\hat{\lambda}_k > \lambda_k/2$ a.s., since if otherwise, then $\left|n^{-1}\hat{\lambda}_k - \lambda_k\right| > \frac{\lambda_k}{2}$ infinitely often with positive probability, and by (27) we have that with positive probability,

$$
\limsup_{n\to\infty} n^{1/2}\log^{-1/2} n \left(\sup_{k>\ell}\left|n^{-1}\hat{\lambda}_k - \lambda_k\right|\right) \geqslant \limsup_{n\to\infty} n^{1/4}\log^{-1/4} n = \infty.
$$

Now, for m large enough,

$$
\begin{aligned}
\left\|\sum_{k=\ell+1}^{m} (n\hat{\lambda}_k^{-1} - \lambda_k^{-1})(\hat{v}_k \otimes \hat{v}_k)\right\| &= \max_{\ell<k\leqslant m}\left|n\hat{\lambda}_k^{-1} - \lambda_k^{-1}\right| \\
&\leqslant \frac{\sup_{k>\ell}\left|n^{-1}\hat{\lambda}_k - \lambda_k\right|}{n^{-1}\hat{\lambda}_m\lambda_m} \\
&\leqslant \frac{2\sup_{k>\ell}\left|n^{-1}\hat{\lambda}_k - \lambda_k\right|}{\lambda_m^2}.
\end{aligned}
$$

By Lemma C.2 and (27) it follows that the above term is $o_p(1)$. Also,

$$
\begin{aligned}
\left\|\sum_{k=\ell+1}^{m} \lambda_k^{-1}(\hat{v}_k \otimes \hat{v}_k - v_k \otimes v_k)\right\| &\leqslant \sum_{k=\ell+1}^{m} \lambda_k^{-1}\left\|\hat{v}_k \otimes (\hat{v}_k - v_k) + (\hat{v}_k - v_k)\otimes v_k\right\| \\
&\leqslant 2\lambda_m^{-1}\sum_{k=\ell+1}^{m}\|\hat{v}_k - v_k\| \qquad (28)\\
&\leqslant 2\lambda_m^{-1}\left\|\bar{\Gamma}_{SS} - \Sigma_{SS}\right\|\sum_{k=\ell+1}^{m}\tau_k.
\end{aligned}
$$

By Lemma C.2 and the assumption of this theorem, we have that the above term is $o_p(1)$.

This completes the proof of (26).

Now write

$$\widehat{A}\widetilde{\Pi}_S - A\widetilde{\Pi}_S = F_1 + F_2 - A(\widetilde{\Pi}_S - \widehat{\Pi}_S).$$

where

$$F_1 = \frac{1}{n}\left(\sum_{t=1}^{n}(\varepsilon_t \otimes f_{t-1})\right)\left(\sum_{k=\ell+1}^{m} n\hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) - \sum_{k=\ell+1}^{m} \lambda_k^{-1}(v_k \otimes v_k)\right)$$

and

$$F_2 = \frac{1}{n}\left(\sum_{t=1}^{n}(\varepsilon_t \otimes f_{t-1})\right)\left(\sum_{k=\ell+1}^{m} \lambda_k^{-1}(v_k \otimes v_k)\right).$$

By (23) and (26), we have that $\|F_1\| = o_p(1)$. Note that

$$F_2 = \frac{1}{n}\left(\sum_{t=1}^{n}(\varepsilon_t \otimes f_{t-1}^S)\right)\Pi_S\left(\sum_{k=\ell+1}^{m} \lambda_k^{-1}(v_k \otimes v_k)\right)$$

and that

$$\left\|\sum_{k=\ell+1}^{m} \lambda_k^{-1}(v_k \otimes v_k)\right\| = \lambda_m^{-1}, \tag{29}$$

by Lemma C.1 and (27), we have that $\|F_2\| = o_p(1)$.

Next, write

$$A\widetilde{\Pi}_S - A\widehat{\Pi}_S = A(\widetilde{\Pi}_S - \Pi_S) + (A\Pi_S - A\underline{\Pi}_S) + A(\underline{\Pi}_S - \widehat{\Pi}_S).$$

Note that $\left\|A(\widetilde{\Pi}_S - \Pi_S)\right\| = o_p(1)$. With a similar argument as in (28), we have that

$$\left\|A(\underline{\Pi}_S - \widehat{\Pi}_S)\right\| = \|A\|\left\|\sum_{k=\ell+1}^{m}\left[(v_k \otimes v_k) - (\hat{v}_k \otimes \hat{v}_k)\right]\right\| = o_p(1).$$

Let $\widetilde{A} = A\Pi_S$. Since A is compact, $\widetilde{A}^*$ is compact. Then $\underline{\Pi}\widetilde{A}^* \to \widetilde{A}^*$ in norm. To see this, write $\underline{\Pi} = \Pi_m$ and notice that if instead $\left\|\underline{\Pi}\widetilde{A}^* - \widetilde{A}^*\right\| \nrightarrow 0$, then there exists $\epsilon > 0$ such that for any n, we may find $x_n \in H$ such that $\|x_n\| = 1$ and that $\left\|(\Pi_{m(n)} - 1)\widetilde{A}^* x_n\right\| > \epsilon$. For any $n' > n$, we have that $\left\|(\Pi_{m(n)} - 1)\widetilde{A}^* x_{n'}\right\| \geqslant \left\|(\Pi_{m(n')} - 1)\widetilde{A}^* x_{n'}\right\| > \epsilon$. Now since $\widetilde{A}^*$ is compact, there exists some subsequence x_{n_i} of x_n such that $\widetilde{A}^* x_{n_i}$ converges in norm to some $x \in H$. Then we have that $\|(\Pi_{m(n)} - 1)x\| > \epsilon$ for all n. However, this is impossible since $m(n) \to \infty$ as $n \to \infty$. Therefore, we have that $\left\|\underline{\Pi}\widetilde{A}^* - \widetilde{A}^*\right\| \to 0$. This then implies that $\left\|\widetilde{A}\underline{\Pi}-\widetilde{A}\right\|\to 0$. That is, $\|A\underline{\Pi}_S - A\Pi_S\|\to 0$. Now we have $\left\|\widehat{A}\widetilde{\Pi}_S - A\widetilde{\Pi}_S\right\| = o_p(1)$, and consistency of $\widehat{A}$ follows immediately.

To obtain the asymptotic distribution of $\widehat{A}$, note that

$$\widehat{A}-\overline{A} = \left(\sum_{t=1}^n \varepsilon_t\otimes f_{t-1}\right)\left(\sum_{k=1}^m \hat{\lambda}_k^{-1}(\hat{v}_k\otimes\hat{v}_k)\right).$$

For any $v\in H_P\equiv H_N$, write

$$n(\widehat{A}-\overline{A})v = G_1+R_1+R_2+R_3,$$

where

$$G_1 = \left(\frac{1}{n}\sum_{t=1}^n \varepsilon_t\otimes f_{t-1}\right)\left(\sum_{k=1}^m n^2\hat{\lambda}_k^{-1}(\hat{v}_k\otimes\hat{v}_k)\right)\widehat{\Pi}_N v,$$

$$R_1 = \left(\frac{1}{n}\sum_{t=1}^n \varepsilon_t\otimes f_{t-1}\right)\left(\sum_{k=1}^\ell n^2\hat{\lambda}_k^{-1}(\hat{v}_k\otimes\hat{v}_k)\right)(\Pi_N-\widehat{\Pi}_N)v,$$

$$R_2 = \left(\sum_{t=1}^n \varepsilon_t\otimes f_{t-1}\right)(\widetilde{\Pi}_S-\Pi_S)\left(\sum_{k=\ell+1}^m n\hat{\lambda}_k^{-1}(\hat{v}_k\otimes\hat{v}_k)\right)(\Pi_N-\widehat{\Pi}_N)v,$$

and

$$R_3 = \left(\sum_{t=1}^n \varepsilon_t\otimes f_{t-1}\right)\Pi_S\left(\sum_{k=\ell+1}^m n\hat{\lambda}_k^{-1}(\hat{v}_k\otimes\hat{v}_k)\right)(\Pi_N-\widehat{\Pi}_N)v.$$

By (23) and Theorem 3.3, we have that $\|R_1\| = O_p(n^{-1})$. By (23), (26), (27) and Theorem 3.3, we have that $\|R_2\| = o_p(n^{-3/4}\log^{-1/4} n)$. Note that $\left(\sum_{t=1}^n \varepsilon_t\otimes f_{t-1}\right)\Pi_S = \sum_{t=1}^n \varepsilon_t\otimes f_{t-1}^S$, by (26), (27), Lemma C.1 and Theorem 3.3, we have that $\|R_3\| = o_p(n^{-1/4}\log^{-1/4} n)$. Now again by (23), (26) and Theorem 3.3 we have that

$$\begin{aligned}
G_1 &= \left(\frac{1}{n}\sum_{t=1}^n \varepsilon_t\otimes f_{t-1}\right)\left(\sum_{k=1}^m n^2\hat{\lambda}_k^{-1}(\hat{v}_k\otimes\hat{v}_k)\right) v \\
&\to_d \left(\int_0^1 (dW\otimes W_P)(r)\right)\left(\sum_{k=1}^\ell \lambda_k^{-1}(v_k\otimes v_k)\right) v \\
&= \left(\int_0^1 (dW\otimes W_P)(r)\right)\left(\int_0^1 (W_P\otimes W_P)(r)dr\right)^+ v,
\end{aligned}$$

where $\int_0^1 (W_p\otimes W_p)(r)dr$ is viewed as an operator restricted on $H_P\equiv H_N$, and $^+$ denote the inverse of the operator on H_P. This then completes the proof for the asymptotic distribution of $\widehat{A}$ on H_N.

For any $v \notin H_N$, we have that

$$(\widehat{A} - \overline{A})v = G_2 + R_4 + R_5 + R_6$$

where

$$G_2 = \frac{1}{\sqrt{n}} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t \otimes f_{t-1}^S \right) \left(\sum_{k=\ell+1}^{m} \lambda_k^{-1} (v_k \otimes v_k) \right) v,$$

$$R_4 = \left(\frac{1}{n} \sum_{t=1}^{n} \varepsilon_t \otimes f_{t-1} \right) \left(\sum_{k=1}^{m} n \hat{\lambda}_k^{-1} (\hat{v}_k \otimes \hat{v}_k) \right) (\Pi_S - \widetilde{\Pi}_S) v,$$

$$R_5 = \left(\frac{1}{n} \sum_{t=1}^{n} \varepsilon_t \otimes f_{t-1} \right) (\widetilde{\Pi}_S - \Pi_S) \left(\sum_{k=1}^{m} n \hat{\lambda}_k^{-1} (\hat{v}_k \otimes \hat{v}_k) \right) \widetilde{\Pi}_S v,$$

and

$$R_6 = \frac{1}{\sqrt{n}} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t \otimes f_{t-1}^S \right) \left(\sum_{k=\ell+1}^{m} \hat{n} \lambda_k^{-1} (\hat{v}_k \otimes \hat{v}_k) - \sum_{k=\ell+1}^{m} \lambda_k^{-1} (v_k \otimes v_k) \right) v.$$

By (23), (24), (26), (27), and Theorem 3.3, we have that $\|R_4\| = o_p(n^{-3/4} \log^{-1/4} n)$. Similarly, $\|R_5\| = o_p(n^{-3/4} \log^{-1/4} n)$. By Lemma C.1 and (26), we have that $\|R_6\| = o_p(n^{-1/2})$.

Note that $Z_t = (\varepsilon_t \otimes f_{t-1}^S) \left(\sum_{k=\ell+1}^{m} \lambda_k^{-1} (v_k \otimes v_k) \right) v = \left\langle f_{t-1}^S, \left(\sum_{k=\ell+1}^{m} \lambda_k^{-1} (v_k \otimes v_k) \right) v \right\rangle \varepsilon_t$ is a martingale difference sequence with respect to $\mathcal{F}_t = \sigma(\varepsilon_i : i \leqslant t)$. Since f_{t-1} is independent of ε_t, we have that

$$\begin{aligned}
\mathbb{E}(Z_t \otimes Z_t) &= \mathbb{E} \left\langle f_{t-1}^S, \left(\sum_{k=\ell+1}^{m} \lambda_k^{-1} (v_k \otimes v_k) \right) v \right\rangle^2 \mathbb{E}(\varepsilon_t \otimes \varepsilon_t) \\
&= \left\langle \left(\sum_{k=\ell+1}^{m} \lambda_k^{-1} (v_k \otimes v_k) \right) v, \mathbb{E}(f_{t-1}^S \otimes f_{t-1}^S) \left(\sum_{k=\ell+1}^{m} \lambda_k^{-1} (v_k \otimes v_k) \right) v \right\rangle \Sigma \\
&= \left\langle \left(\sum_{k=\ell+1}^{m} \lambda_k^{-1} (v_k \otimes v_k) \right) v, \left(\sum_{k=\ell+1}^{m} (v_k \otimes v_k) \right) v \right\rangle \Sigma \\
&= \left(\sum_{k=\ell+1}^{m} \lambda_k^{-1} \langle v_k, v \rangle^2 \right) \Sigma.
\end{aligned}$$

By the central limit theorem for real-valued martingale difference sequence, we have that for any $x \in H$,

$$\frac{1}{s_m(v)\sqrt{n}} \sum_{t=1}^{n} \langle x, Z_t \rangle \to_d \mathbb{N}(0, \langle x, \Sigma x \rangle) \tag{30}$$

where $s_m^2(x) = \sum_{k=\ell+1}^m \lambda_k^{-1}\langle v_k, v\rangle^2$. Next, we show that the sequence $\widetilde{Z}_n = \frac{1}{s_m(v)\sqrt{n}}\sum_{t=1}^n Z_t$ is tight. Let Π_n^Σ be the orthogonal projection onto the space spanned by the first n eigenfunctions of the variance operator Σ. Since that for any $\epsilon > 0$, $\mathbb{P}\left(\left\|(1-\Pi_n^\Sigma)\widetilde{Z}_n\right\| > \epsilon\right) \leqslant \frac{\mathbb{E}\left\|(1-\Pi_n^\Sigma)\widetilde{Z}_n\right\|^2}{\epsilon^2}$ and that

$$
\begin{aligned}
\mathbb{E}\left\|(1-\Pi_n^\Sigma)\widetilde{Z}_n\right\|^2 &= \frac{1}{ns_m^2(v)}\mathbb{E}\left(\left\|\sum_{t=1}^n \left\langle \left(\sum_{k=\ell+1}^m \lambda_k^{-1}(v_k\otimes v_k)\right) f_{t-1}^S, v\right\rangle (1-\Pi_n^\Sigma)\varepsilon_t\right\|^2\right) \\
&= \frac{1}{ns_m^2(v)}\sum_{t=1}^n \mathbb{E}\left(\left\langle \left(\sum_{k=\ell+1}^m \lambda_k^{-1}(v_k\otimes v_k)\right) f_{t-1}^S, v\right\rangle^2 \left\|(1-\Pi_n^\Sigma)\varepsilon_t\right\|^2\right) \\
&= \frac{1}{s_m^2(v)}\mathbb{E}\left\|(1-\Pi_n^\Sigma)\varepsilon_t\right\|^2 \mathbb{E}\left\langle \left(\sum_{k=\ell+1}^m \lambda_k^{-1}(v_k\otimes v_k)\right) f_{t-1}^S, v\right\rangle^2 \\
&= \operatorname{tr}\left((1-\Pi_n^\Sigma)\Sigma\right) \to 0
\end{aligned}
$$

as $n\to\infty$, we have

$$\limsup_{n\to\infty}\sup_n \mathbb{P}\left(\left\|(1-\Pi_n^\Sigma)\widetilde{Z}_n\right\| > \epsilon\right) = 0.$$

This implies that $(\widetilde{Z}_t)$ is tight, so the central limit theorem for the real valued martingale difference sequence as in (30) implies a central limit limit theorem for the H-valued martingale difference sequence Z_t:

$$\frac{\sqrt{n}}{s_m(x)}G_2 = \frac{1}{s_m(v)\sqrt{n}}\sum_{t=1}^n Z_t \to_d \mathbb{N}(0,\Sigma).$$

This completes the proof for the asymptotic distribution of $\widehat{A}$ outside H_N. $\square$

Proof of Corollary 3.5. Since

$$\left\|(\overline{A}-A)\Pi_N\right\| \leqslant \left\|A(\widehat{\Pi}-\underline{\Pi})\Pi_N\right\| + \left\|A(\underline{\Pi}-1)\Pi_N\right\| \leqslant \|A\|\left\|\widehat{\Pi}-\underline{\Pi}\right\|,$$

it then suffices to show that

$$\left\|\widehat{\Pi}-\underline{\Pi}\right\| = o_p(n^{-1/2}m^{1/2}). \tag{31}$$

Write $\left\|\widehat{\Pi}-\underline{\Pi}\right\| = \left\|\widehat{\Pi}_N + \widehat{\Pi}_S - \Pi_N - \underline{\Pi}_S\right\| \leqslant \left\|\widehat{\Pi}_N - \Pi_N\right\| + \left\|\widehat{\Pi}_S - \underline{\Pi}_S\right\|$. Since $\left\|\widehat{\Pi}_N - \Pi_N\right\| = O_p(n^{-1})$, it suffices to show that $\left\|\widehat{\Pi}_S - \underline{\Pi}_S\right\| = o_p(n^{-1/2}m^{1/2})$.

Let $(\overline{v}_k^S)$ be the eigenfunctions associated with the nonzero eigenvalues $(\overline{\lambda}_k^S)$ of $\overline{\Gamma}_{SS}$ such that $\overline{\lambda}_1^S \geqslant \overline{\lambda}_2^S \geqslant \cdots$, and define $\overline{\Pi}_S$ to be the orthogonal projection on the subspace of H spanned by $(\overline{v}_k^S)$ for $k = 1, \ldots, m-\ell$. Write

$$\left\|\widehat{\Pi}_S - \underline{\Pi}_S\right\| \leqslant \left\|\widehat{\Pi}_S - \overline{\Pi}_S\right\| + \left\|\overline{\Pi}_S - \underline{\Pi}_S\right\|. \tag{32}$$

For the first term on the right hand side of (32), it follows from Hu, Park and Qian (2016) that under Assumption 3.2, $\left\|\widehat{\Pi}_S - \overline{\Pi}_S\right\| = o_p(n^{-1/2}m^{1/2})$. To analyze the second term in (32), we note that $\overline{\Pi}_S$ and $\underline{\Pi}_S$ are orthogonal projections onto the subspaces spanned by the leading eigenvalues of $\widetilde{\Pi}_S\widehat{\Gamma}\widetilde{\Pi}_S$ and of $\Pi_S\widehat{\Gamma}\Pi_S$, respectively. It follows from Lemma 3.3 that $\widetilde{\Pi}_S\Pi_N = O_p(n^{-1})$ and $\Pi_N\widetilde{\Pi}_S = O_p(n^{-1})$, then we may derive that $\widetilde{\Pi}_S\widehat{\Gamma}\widetilde{\Pi}_S = \overline{\Gamma}_{SS} + O_p(n^{-1})$ uniformly in n, and consequently, $\max_{1\leqslant k\leqslant m-\ell}\left\|\hat{v}_{\ell+k} - \overline{v}_k^S\right\| = O_p(n^{-1})$, from which it follows that $\left\|\overline{\Pi}_S - \underline{\Pi}_S\right\| = O_p(n^{-1}m)$. Now it follows that $\left\|\widehat{\Pi}_S - \underline{\Pi}_S\right\| = o_p(n^{-1/2}m^{1/2})$, which shows that $(\overline{A} - A)\Pi_N = o_p(n^{-1/2}m^{1/2})$.

Also,

$$\begin{aligned}\left\|(\overline{A} - A)\Pi_S v\right\| &\leqslant \left\|A(\widehat{\Pi} - \underline{\Pi})\Pi_S v\right\| + \left\|A(\underline{\Pi} - 1)\Pi_S v\right\| \\ &\leqslant \|A\|\left\|\widehat{\Pi} - \underline{\Pi}\right\|\|v\| + \|A\|\left\|(\underline{\Pi} - 1)v\right\| \\ &= o_p(n^{-1/2}m^{1/2}) + O(\|(1 - \underline{\Pi})v\|),\end{aligned}$$

and the proof is complete. □

Proof of Theorem 3.7. Write

$$\widetilde{B} - \overline{B}_S = \left(\sum_{t=1}^n \left[\varepsilon_t + (A-1)\widehat{f}_{t-1}^N\right] \otimes \widehat{f}_{t-1}^S\right)\left(\sum_{t=1}^n \widehat{f}_{t-1}^S \otimes \widehat{f}_{t-1}^S\right)^+.$$

Note that $\varepsilon_t + (A-1)\widehat{f}_{t-1}^N = \varepsilon_t + (A-1)(\widehat{\Pi}_N - \Pi_N)f_{t-1} = \varepsilon_t + O_p(n^{-1/2})$.

Since

$$\begin{aligned}&\left(\sum_{t=1}^n \left[\varepsilon_t + (A-1)\widehat{f}_{t-1}^N\right] \otimes \widehat{f}_{t-1}^S\right)\left(\sum_{t=1}^n \widehat{f}_{t-1}^S \otimes \widehat{f}_{t-1}^S\right)^+ \Pi_N \\ &= \left(\sum_{t=1}^n \left[\varepsilon_t + (A-1)\widehat{f}_{t-1}^N\right] \otimes \widehat{f}_{t-1}^S\right)\left(\sum_{t=1}^n \widehat{f}_{t-1}^S \otimes \widehat{f}_{t-1}^S\right)^+ \left(\Pi_N - \widehat{\Pi}_N\right) \\ &= \left(\left(\frac{1}{n}\sum_{t=1}^n \varepsilon_t \otimes f_{t-1}^S\right)\left(\sum_{k=\ell+1}^m \lambda_k^{-1}(v_k \otimes v_k)\right) + o_p(1)\right)\left(\Pi_N - \widehat{\Pi}_N\right),\end{aligned}$$

by (27) and (29), the above term is $o_p(n^{-5/4}\log^{-1/4} n)$. Therefore, for any $v \in H_N$, $(\widetilde{B} -$

$\overline{B}_S)v = o_p(n^{-1})$.

Similarly, for any $v \notin H_N$, we may show that

$$\frac{\sqrt{n}}{s_m(v)}\left(\widetilde{B}-\overline{B}_S\right)v = \frac{\sqrt{n}}{s_m(v)}\left(\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t \otimes f_{t-1}^S\right)\left(\sum_{k=\ell+1}^{m}\lambda_k^{-1}(v_k\otimes v_k)\right)v + o_p(1).$$

It then follows from the proof of Theorem 3.4 that $\frac{\sqrt{n}}{s_m(v)}\left(\widetilde{B}-\overline{B}_S\right)v \to_d \mathbb{N}(0,\Sigma)$ for any $v \notin H_N$.

Consistency of $\widetilde{A}$ follows from the above results easily. □

Proof of Theorem 3.8. We first prove the consistency of $\widehat{\Pi}_P$. Consider the operator

$$B = \underline{\Pi} + (\mathbb{E}\Delta(\underline{\Pi}f_t)\otimes(\underline{\Pi}_S f_{t-1}))\,(\mathbb{E}(\underline{\Pi}_S f_{t-1})\otimes(\underline{\Pi}_S f_{t-1}))^{+}$$

where $^+$ denotes the inverse on $\underline{H}_S$. It is obvious that B has kernel $\underline{H}^\perp$, and has an invariant subspace H_P that corresponds to the eigenvalue 1. Since $\mathbb{E}\Delta(\underline{\Pi}f_t)\otimes(\underline{\Pi}_S f_{t-1}) = \mathbb{E}\Delta(\underline{\Pi}f_t^T)\otimes(\underline{\Pi}_S f_{t-1})$, we see that $\underline{\Pi}H_T = H_T \cap \underline{H}$ is an invariant subspace of B. Note that B is in fact essentially an operator restricted on the finite dimensional subspace $\underline{H}$, in view of the comments in Section 2.3, the projection with range $H_T \cap \underline{H}$ and kernel space $H_P \oplus \underline{H}^\perp$ is given by

$$\begin{aligned}(B-\underline{\Pi})(\underline{\Pi}_S(B-\underline{\Pi}))^{+}\underline{\Pi}_S &= \Big(\mathbb{E}\Delta(\underline{\Pi}f_t)\otimes(\underline{\Pi}_S f_{t-1})\Big)\Big(\mathbb{E}\Delta(\underline{\Pi}_S f_t)\otimes(\underline{\Pi}_S f_{t-1})\Big)^{+}\\ &= \underline{\Pi}_S + \Big(\mathbb{E}\Delta(\Pi_N f_t)\otimes(\underline{\Pi}_S f_{t-1})\Big)\Big(\mathbb{E}\Delta(\underline{\Pi}_S f_t)\otimes(\underline{\Pi}_S f_{t-1})\Big)^{+}.\end{aligned}$$

The projection with range H_P and kernel space $(H_T \cap \underline{H})\oplus\underline{H}^\perp$ is therefore $\Pi_P^\circ = \Pi_N - (\mathbb{E}\Delta(\Pi_N f_t)\otimes(\underline{\Pi}_S f_{t-1}))(\mathbb{E}\Delta(\underline{\Pi}_S f_{t-1})\otimes(\underline{\Pi}_S f_{t-1}))^{+}$ since the two projections should add up to $\underline{\Pi}$.

The sample analog of the latter projection is obviously $\widehat{\Pi}_P$. We next show that $\widehat{\Pi}_P - \Pi_P^\circ = o_p(1)$. First write

$$\frac{1}{n}\sum_{t=1}^{n}(\Delta\widehat{f}_t^S \otimes \widehat{f}_{t-1}^S) = G_1 + R_1 + R_2$$

where

$$G_1 = \frac{1}{n}\sum_{t=1}^{n}\underline{\Pi}_S(\Delta f_t^S \otimes f_{t-1}^S)\underline{\Pi}_S,$$

$$R_1 = (\widehat{\Pi}_S - \underline{\Pi}_S) \left(\frac{1}{n} \sum_{t=1}^{n} (\Delta f_t \otimes f_{t-1}) \right) \widehat{\Pi}_S,$$

and

$$R_2 = \underline{\Pi}_S \left(\frac{1}{n} \sum_{t=1}^{n} (\Delta f_t \otimes f_{t-1}) \right) (\widehat{\Pi}_S - \underline{\Pi}_S).$$

It follows from Lemma C.1 and (31) that R_1 and R_2 are both $o_p(1)$ terms. Note that $G_1 = \mathbb{E}\Delta(\underline{\Pi}_S f_t) \otimes (\underline{\Pi}_S f_{t-1}) + o_p(1)$, we therefore have that $\frac{1}{n}\sum_{t=1}^{n}(\Delta \widehat{f}_t^S \otimes \widehat{f}_{t-1}^S) = \mathbb{E}\Delta(\underline{\Pi}_S f_t) \otimes (\underline{\Pi}_S f_{t-1}) + o_p(1)$. Next, write

$$\frac{1}{n} \sum_{t=1}^{n} (\Delta \widehat{f}_t^N \otimes \widehat{f}_{t-1}^S) = G_2 + R_3 + R_4$$

where

$$G_2 = \frac{1}{n} \sum_{t=1}^{n} (\varepsilon_t^N \otimes f_{t-1}^S) \underline{\Pi}_S,$$

$$R_3 = (\widehat{\Pi}_N - \Pi_N) \left(\frac{1}{n} \sum_{t=1}^{n} (\Delta f_t \otimes f_{t-1}) \right) \widehat{\Pi}_S,$$

and

$$\begin{aligned} R_4 &= \left(\frac{1}{n} \sum_{t=1}^{n} (\varepsilon_t^N \otimes f_{t-1}) \right) (\widehat{\Pi}_S - \underline{\Pi}_S) \\ &= \left(\frac{1}{n} \sum_{t=1}^{n} (\varepsilon_t^N \otimes f_{t-1}^S) \right) (\widehat{\Pi}_S - \underline{\Pi}_S) + \left(\frac{1}{n} \sum_{t=1}^{n} (\varepsilon_t^N \otimes f_{t-1}^N)(\Pi_N - \widehat{\Pi}_N) \right) \widehat{\Pi}_S. \end{aligned}$$

By Lemma C.1 and (31) we have that $R_3 = O_p(n^{-1}), R_4 = o_p(n^{-1}m^{1/2})$, and $G_2 = \mathbb{E}\Delta(\Pi_N f_t) \otimes (\underline{\Pi}_S f_{t-1}) + O_p(n^{-1/2})$. It then follows that $\frac{1}{n}\sum_{t=1}^{n}(\Delta \widehat{f}_t^N \otimes \widehat{f}_{t-1}^S) = \mathbb{E}\Delta(\underline{\Pi}_S f_t) \otimes (\underline{\Pi}_S f_{t-1}) + O_p(n^{-1/2})$.

Since

$$\begin{aligned} \mathbb{E}\Delta(\underline{\Pi}_S f_t) \otimes (\underline{\Pi}_S f_{t-1}) &= (\underline{\Pi}_S A \underline{\Pi}_S - \underline{\Pi}_S) \mathbb{E} \underline{\Pi}_S (f_{t-1} \otimes f_{t-1}) \underline{\Pi}_S + o(1) \\ &= (\underline{\Pi}_S A \underline{\Pi}_S - \underline{\Pi}_S) \sum_{k=\ell+1}^{m} \lambda_k (v_k \otimes v_k) + o(1), \end{aligned}$$

by (27) and (29) we have that $\mathbb{E}\Delta(\underline{\Pi}_S f_t) \otimes (\underline{\Pi}_S f_{t-1}) = o(n^{1/4} \log^{-1/4} n)$. It then follows that

$$\widehat{\Pi}_P = \widehat{\Pi}_N + \Big(\mathbb{E}\Delta(\Pi_N f_t) \otimes (\underline{\Pi}_S f_{t-1}) + O_p(n^{-1/2})\Big)\Big(\mathbb{E}\Delta(\underline{\Pi}_S f_t) \otimes (\underline{\Pi}_S f_{t-1}) + o_p(1)\Big)$$
$$= \Pi_P^\circ + o_p(1).$$

Following similar steps, we can show that $\widetilde{A} - B = o_p(1)$. By consistency of $\widetilde{A}$, we have that $B - A = o(1)$. Since Π_P° is the eigen-projection of B corresponding to the eigenvalue 1, and Π_P is the eigen-projection of A corresponding to the eigenvalue 1, and that $B \to A$, we have that $\Pi_P^\circ - \Pi_P = o_p(1)$. Since $\widehat{\Pi}_P - \Pi_P^\circ = o_p(1)$, we have that $\widehat{\Pi}_P - \Pi_P = o_p(1)$.

Since $\Pi_P + \Pi_T = \widehat{\Pi}_P + \widetilde{\Pi}_T = 1$, we have $\widetilde{\Pi}_T - \Pi_T = o_p(1)$. □

Proof of Lemma 3.10. Write

$$\widehat{A} f_n - \overline{A} f_n = G + R_1 + R_2$$

where

$$R_1 = (\widehat{A} - \overline{A})\widehat{\Pi}_N f_n,$$
$$R_2 = (\widehat{A} - \overline{A})(\widetilde{\Pi}_S - \Pi_S) f_n,$$

and

$$G = (\widehat{A} - \overline{A})\Pi_S f_n.$$

By Lemma 3.1 and (25) we have that $\|R_1\| = O_p(n^{-1/2})$. Since $\|R_2\| \leqslant \left\|\widehat{A} - \overline{A}\right\| \left\|\widetilde{\Pi}_S - \Pi_S\right\| \|f_n\|$, by Lemma 3.1 and the consistency of $\hat{A}$, we have that $\|R_2\| = o_p(n^{-1/2})$.

Following the proof of Theorem 3.4, we have that $G = \widetilde{G} + o_p(n^{-1/2})$ where

$$\widetilde{G} = \frac{1}{n}\left(\sum_{t=1}^{n}(\varepsilon_t \otimes f_{t-1}^S)\right)\left(\sum_{k=\ell+1}^{m}\lambda_i^{-1}(v_i \otimes v_i)\right) f_n^S.$$

Therefore, it suffices to show that

$$\sqrt{n/m}\,\widetilde{G} \to_d \mathbb{N}(0, \Sigma). \tag{33}$$

We follow Mas (2007) for this proof. Specifically, we follow its convention to show that

$$\sqrt{\frac{1}{nm}}\left(\sum_{t=1}^{n}(\varepsilon_t \otimes f_{t-1}^S)\right)\left(\sum_{k=\ell+1}^{m}\lambda_k^{-1}(v_k \otimes v_k)\right) f_{n+1}^S \to_d \mathbb{N}(0, \Sigma).$$

This is obviously equivalent to (33) if $\widehat{A}$ is estimated using data only up to time $n-1$. For convenience, we write $Q_2^+ = \sum_{k=\ell+1}^m \lambda_k^{-1}(v_k \otimes v_k)$. Since A restricted on H_N is the identity operator, we have that

$$f_t^S = \Pi_S(Af_{t-1} + \varepsilon_t) = \Pi_S(A(f_{t-1}^N + f_{t-1}^S)) + \varepsilon_t^S = \Pi_S A \Pi_S f_{t-1}^S + \varepsilon_t^S.$$

This implies that (f_t^S) has a functional autoregressive representation with autoregressive operator $\Pi_S A \Pi_S$. For convenience, let $A_S = \Pi_S A \Pi_S$. Since $\|A_S\| \leqslant \|A\|$, the first order difference equation $g_t = A_S g_{t-1} + \varepsilon_t^S$ has a unique stationary solution. Since $g_t = f_t^S$ is a solution, it is the only solution. This implies that we may view (f_t^S) as a stationary functional autoregressive process by itself.

Now write

$$\begin{aligned}
&\left(\sum_{t=1}^n (\varepsilon_t \otimes f_{t-1}^S)\right)\left(\sum_{k=\ell+1}^m \lambda_i^{-1}(v_i \otimes v_i)\right) f_{n+1}^S \\
&= \sum_{t=1}^n \left\langle f_{t-1}^S, Q_2^+ f_{n+1}^S \right\rangle \varepsilon_t \\
&= \sum_{t=1}^n \left\langle Q_2^+ f_{t-1}^S, f_{n+1}^S \right\rangle \varepsilon_t \\
&= \sum_{t=1}^n (Z_t^+ + Z_t^0 + Z_t^-)
\end{aligned}$$

where

$$\begin{aligned}
Z_t^+ &= \left\langle Q_2^+ f_{t-1}^S, f_{t+}^S \right\rangle \varepsilon_t, \\
Z_t^0 &= \left\langle Q_2^+ f_{t-1}^S, (A_S)^{n+1-t} \varepsilon_t^S \right\rangle \varepsilon_t, \\
Z_t^- &= \left\langle Q_2^+ f_{t-1}^S, (A_S)^{n+2-t} f_{t-1}^S \right\rangle \varepsilon_t,
\end{aligned}$$

and

$$f_{t+}^S = \varepsilon_{n+1}^S + A_S \varepsilon_n^S + \cdots + (A_S)^{n-t} \varepsilon_{t+1}^S.$$

Minor modifications of the proof of Lemma 5.7 in Mas (2007) shows that (Z_t^+) and (Z_t^-) are H-valued martingale difference sequences with respect to $\mathcal{F}_t$. Following the proof of Lemma 5.8 in Mas (2007), it is easy to show that $\mathbb{E}(Z_t^+ \otimes Z_s^+) = 0$ for $t < s$, and $\mathbb{E}(Z_t^+ \otimes Z_t^+) = \mathbb{E}\left\langle Q_2^+ f_{t-1}^S, f_{t+}^S \right\rangle^2 \Sigma$. Note that Q_2^+ is the inverse of $\mathbb{E}(f_{t-1}^S \otimes f_{t-1}^S)$ restricted on $\underline{H}_S$, we may follow the proof of Lemma 5.8 in Mas (2007) to obtain that

$$\mathbb{E}\left(Z_t^+ \otimes Z_t^+\right) = \left(m - \ell - \operatorname{tr}\left(Q_2^+ (A_S)^{n-t+1} \Gamma_{SS} (A_S^*)^{n-t+1}\right)\right) \Sigma.$$

Since $\left|\mathrm{tr}\left(Q_2^+(A_S)^{n-t+1}\Gamma_{SS}(A_S^*)^{n-t+1}\right)\right|$ is bounded by a constant under the assumption in the theorem, we have that $\mathbb{E}\left\langle Q_2^+ f_{t-1}^S, f_{t+}^S\right\rangle^2 = O(nm)$. Then since $(\langle Z_t^+, x\rangle)$ is a martingale difference sequence for any $x \in H$, by the central limit theorem we have that

$$\frac{1}{\sqrt{nm}}\sum_{t=1}^{n}\langle Z_t^+, x\rangle \to_d \mathbb{N}(0, \langle x, \Sigma x\rangle).$$

We may follow the proof of Lemma 5.9 in Mas (2007) to show that $\frac{1}{\sqrt{nm}}\sum_{t=1}^n Z_t^+$ is a tight sequence. This then implies that $\frac{1}{\sqrt{nm}}\sum_{t=1}^n Z_t^+ \to_d \mathbb{N}(0, \Sigma)$. One may follow Lemma 5.10 in Mas (2007) to show that $\frac{1}{\sqrt{nm}}\sum_{t=1}^n Z_t^0 \to_p 0$ and that $\frac{1}{\sqrt{nm}}\sum_{t=1}^n Z_t^- \to_p 0$. The conclusion then follows immediately. $\square$

Proof of Theorem 3.11. In view of Lemma 3.10, it suffices to show that $\sqrt{n/m}(\overline{A}f_n - Af_n) = o_p(1)$. Notice that

$$\overline{A}f_n - Af_n = (\overline{A}\widehat{\Pi}_N f_n - A\widehat{\Pi}_N f_n) + (\overline{A}\widetilde{\Pi}_S f_n - A\widetilde{\Pi}_S f_n) = \overline{A}\widetilde{\Pi}_S f_n - A\widetilde{\Pi}_S f_n,$$

it then suffices to show that $\sqrt{n/m}(\overline{A}\widetilde{\Pi}_S f_n - A\widetilde{\Pi}_S f_n) = o_p(1)$. Since

$$\overline{A}\widetilde{\Pi}_S f_n - A\widetilde{\Pi}_S f_n = (\overline{A}\Pi_S - A\Pi_S)f_n + (\overline{A} - A)(\widetilde{\Pi}_S - \Pi_S)f_n,$$

and that $\left\|\overline{A} - A\right\| \leqslant \|A\|$, by Lemma 3.1, it suffices to show that

$$\left\|(\overline{A}\Pi_S - A\Pi_S)f_n\right\| = \left\|(\overline{A} - A)f_n^S\right\| = o_p(n^{-1/2}m^{1/2}).$$

Write

$$(\overline{A} - A)f_n^S = A(\widehat{\Pi} - \underline{\Pi})f_n^S + A(\underline{\Pi} - 1)f_n^S.$$

In the proof of Corollary 3.5, we have shown that $\left\|\widehat{\Pi} - \underline{\Pi}\right\| = o_p(n^{-1/2}m^{1/2})$ under Assumption 3.2. Also,

$$\mathbb{E}\left\|(\underline{\Pi} - 1)f_n^S\right\|^2 = \mathbb{E}\left\|\sum_{k=m+1}^{\infty}\langle v_k, f_n^S\rangle v_k\right\|^2 = \mathbb{E}\left(\sum_{k=m+1}^{\infty}\langle v_k, f_n\rangle^2\right) = \sum_{k=m+1}^{\infty}\lambda_k = o(n^{-1}m)$$

by Assumption 3.4, which implies that $\left\|(\underline{\Pi} - 1)f_n^S\right\| = o_p(n^{-1/2}m^{1/2})$. This then completes the proof. $\square$

References

Bosq, D. (2000): *Linear Processes in Function Spaces: Theory and Applications*, Lecture Notes in Statistics 149, New York: Springer.

Chang, Y., B. Hu, and J. Y. Park (2016a): "A Study of Distributional Earning Dynamics," Manuscript.

Chang, Y., R. Kaufmann, C. S. Kim, J. I. Miller, S. Park, and J. Y. Park (2020): "Evaluating Trends in Time Series of Distributions: A Spatial Fingerprint of Human Effects on Climate," *Journal of Econometrics*, 214, 274–294.

Chang, Y., C. S. Kim, and J. Y. Park (2016b): "Nonstationarity in time series of state densities," *Journal of Econometrics*, 192, 152–167.

Chen, Y. and B. Li (2016): "An Adaptive Functional Autoregressive Forecast Model to Predict Electricity Price Curves," *Forthcoming in Journal of Business & Economic Statistics*.

Ferraty, F. and P. Vieu (2006): *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Series in Statistics, New York: Springer.

Gohberg, I., S. Goldberg, and M. A. Kaashoek (1990): *Classes of Linear Operators*, vol. I, Basel: Birkhäuser.

Gürkaynak, R., B. Sack, and J. Wright (2007): "The U.S. Treasury Yield Curve: 1961 to the Present," *Journal of Monetary Economics*, 54, 2291–2304.

Halmos, P. R. (1974): *Finite Dimensional Vector Spaces*, New York: Springer-Verlag.

Hays, S., H. Shen, and J. Z. Huang (2012): "Functional Dynamic Factor Models with Application to Yield Curve Forecasting," *The Annals of Applied Statistics*, 6, 870–894.

Horväth, L. and P. Kokoszka (2012): *Inference for Functional Data with Applications*, Springer Series in Statistics, New York: Springer.

Hu, B., J. Y. Park, and J. Qian (2016): "Analysis of Distributional Dynamics for Repeated Cross-Sectional and Intra-Period Observations," Manuscript.

Kargin, V. and A. Onatski (2008): "Curve Forecasting by Functional Autoregression," *Journal of Multivariate Analysis*, 99, 2508–2526.

Kneip, A. and K. J. Utikal (2001): "Inference for Density Families Using Functional Principal Component Analysis," *Journal of the American Statistical Association*, 96, 519–532.

Kuelbs, J. (1973): "The Invariance Principle for Banach Space Valued Random Variables," *Journal of Multivariate Analysis*, 3, 162–172.

Mas, A. (2007): "Weak Convergence in the Functional Autoregressive Model," *Journal of Multivariate Analysis*, 98, 1231–1261.

McCracken, M. and S. Ng (2016): "FRED-MD: A Monthly Database for Macroeconomic Research," *Journal of Business & Economic Statistics*, 34, 574–589.

Miranda-Agrippino, S. and G. Ricco (2021): "The transmission of monetary policy shocks," *American Economic Journal: Macroeconomics*, 13, 74–107.

Park, J. Y. and P. Phillips (1988): "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory*, 4, 468–497.

——— (1989): "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory*, 5, 95–131.

Park, J. Y. and J. Qian (2012): "Functional Regression of Continuous State Distributions," *Journal of Econometrics*, 167, 397–412.

Ramsey, J. O. and B. W. Silverman (2005): *Functional Data Analysis*, Springer Series in Statistics, New York: Springer-Verlag, 2nd ed.

Romer, C. D. and D. H. Romer (2004): "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94, 1055–1084.

——— (2010): "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review*, 100, 763–801.

Velasco, M. V. and A. R. Villena (1995): "A Random Banach-Steinhaus Thoerem," *Proceedings of the American Mathematical Society*, 123, 2489–2497.