Econometric Analysis of Functional Dynamics in the Presence of Persistence

Bo Hu†
Department of Economics
Indiana University

Joon Y. Park‡
Department of Economics
Indiana University
and Sungkyunkwan University

March 27, 2017

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 $T$, 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 $\sqrt{T}$ 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 Beveridge-Nelson decomposition, function forecasting

*We are grateful for useful comments to Yoosoon Chang, Juan Carlos Escanciano, the participants of 2016 Midwest Econometrics Group Meeting, and the seminar attendants at Indiana University.

†Email: hu21@indiana.edu
‡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., 2016c), 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., 2016b).

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. Horváth 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. (2016d) 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 an estimator for the functional autoregressive operator based on functional principal component analysis. Our estimator is consistent under very mild regularity conditions, and converges at different rates on different subspaces. In the nonstationary subspace, our estimator converges at rate $T$, and the limit distribution is nonstandard, given as a function of Brownian motions. Elsewhere, our estimator converges at the parametric $\sqrt{T}$-rate or at a rate slower than $\sqrt{T}$, depending on the subspaces in which the convergence is considered, and the limit distributions are Gaussian. 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 $\sqrt{T}$. 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 United States government bond yields. We model the time series of the forward rate curves by a functional autoregressive model with two unit roots. We find that the future short term, medium term and long term interest rates respond to changes in the current forward rates differently. The future short term interest rate responds mainly to changes in the current forward rates for short maturities, while the future long term interest rate responds to changes in the current forward rates for almost all maturities. Furthermore, we find evidence that the current expected interest rates predict future spot rates, which lends support to the rational expectation theory. In addition, we separate the permanent component from the transitory component in the time series of forward rate curves, and propose factors for both components. Using the adaptive least absolute shrinkage and selection operator (LASSO), we find that one of our permanent factors is related to the long rate, and our stationary factors are related to the short rate, the foreign exchange market, the housing market, the labor market and the volatility index.

Finally, we conduct simulations to evaluate the performance of our estimation and prediction procedures. We compare our predictor with two other benchmark predictors which are expected to perform best for nonstationary martingale processes and for iid processes.
We find that our FAR predictor outperforms its competitors since it accommodates both the stationary and the nonstationary dynamics in the process.

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.

2 Model and Background

2.1 The Model

Throughout 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 \((\Omega, F, \mathbb{P})\), i.e., \(f_t : \Omega \rightarrow H\), for each \(t = 1, 2, \ldots\). A frequently used space for \(H\) in economic applications is the space \(L^2(C)\) of all square integrable real-valued functions on some subset \(C\) of the set \(\mathbb{R}\) of real numbers. For example, Kneip and Utikal (2001) model the density functions in the space \(L^2(\mathbb{R})\) and Kargin and Onatski (2008) analyze the Eurodollar futures rate curves in the space \(L^2([0,1])\). Hu et al. (2016) study the dynamics of the demeaned density functions, which belongs to a subspace of \(L^2(C)\) consisting of all functions integrated to zero and defined on a compact subset \(C\) of \(\mathbb{R}\). In what follows, we consider \(H\) as a Hilbert space over the set \(\mathbb{C}\) of complex numbers, and denote by \(\langle \cdot, \cdot \rangle\) and \(\|\cdot\|\) the inner product and the norm defined in \(H\) respectively. In particular, when \(H = L^2(\mathbb{C})\), we define \(\langle u, v \rangle = \int u(x)\overline{v(x)}dx\).

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

\[
 f_t = Af_{t-1} + \varepsilon_t, 
\]

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

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 $\langle v, \mathbb{E} f \rangle = \mathbb{E} \langle v, f \rangle$. Moreover, for any $H$-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 $H$-valued random variable $f$ such that $\mathbb{E} \|f\|^2 < \infty$. Loosely put, we may view $\mathbb{E} f$ as the expectation of an infinite dimensional vector, and $\mathbb{E} (f \otimes f)$ as the expectation of the outer product of two infinite dimensional random vectors. An $H$-valued white noise ($\varepsilon_t$) is a stochastic process such that $\mathbb{E} \varepsilon_t = 0$ for all $t$, $\mathbb{E} (\varepsilon_t \otimes \varepsilon_t)$ is independent of $t$, and $\mathbb{E} (\varepsilon_t \otimes \varepsilon_s) = 0$ for $t \neq s$.

If $\|A^r\| < 1$ for some $r \in \mathbb{N}$, 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. (2016d) 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 $\lambda$ 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) $(\varepsilon_t)$ is independent and identically distributed with mean zero and covariance operator $\Sigma$, is independent of $f_0$, and $E \|\varepsilon_t\|^4 < \infty$.

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. 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 $(\varepsilon_t)$ being independent and identically distributed with $E \|\varepsilon_t\|^4 < \infty$ is made for simplicity, and we may readily allow $(\varepsilon_t)$ to be a general martingale difference sequence with $\sup_{t \geq 1} E(\|\varepsilon_t\|^{2+\epsilon} | \mathcal{F}_{t-1}) < \infty$ 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

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

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-$\delta$ function with a spike at $x$, where the superscript $^*$ denotes the adjoint. Similarly, $A^{*t}v$ may be viewed as the response function of $\langle v, f_t \rangle$ to impulses to $f_{t-i}$.

---

1 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 $A$. 
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 \(\Gamma_P\) and \(\Gamma_T\) be two non-intersecting Cauchy contours on the complex plane such that 1 lies in the inner domain of \(\Gamma_P\) and \(\lambda(A)\setminus\{1\}\) lies in the inner domain of \(\Gamma_T\). 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 \(\Pi_P\) and \(\Pi_T\) are independent of the choices of \(\Gamma_P\) and \(\Gamma_T\). Finally, we denote the images of \(\Pi_P\) and \(\Pi_T\) 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

\[ f_t = f^P_t + f^T_t, \tag{3} \]

where

\[ f^P_t = \Pi_P f_t \quad \text{and} \quad f^T_t = \Pi_T f_t, \]

and similarly, \(\varepsilon_t = \varepsilon^P_t + \varepsilon^T_t\) with \(\varepsilon^P_t = \Pi_P \varepsilon_t\) 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 \(Af^P_t \in H_P\) and \(Af^T_t \in H_T\), and therefore, we may easily deduce that

\[ f^P_t = A f^P_{t-1} + \varepsilon^P_t \tag{4} \]
Figure 1: Decomposition of Functional Time Series

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

and

\[ f^T_t = A_T f^T_{t-1} + \varepsilon^T_t \]  \hspace{1cm} (5)

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 \((f^P_t)\) 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 \(m\)-dimensional, and \(A_P\) be a linear transformation on an \(m\)-dimensional vector space. Therefore, 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 \(1 \leq d \leq m\). 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 \((f^P_t)\).

**Lemma 2.2.** \(A_P - 1\) is nilpotent of degree \(d\) if and only if \((f^P_t)\) 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 \((f^P_t)\) becomes a random walk. Moreover, we let \(\|A_T^r\| < 1\) for some \(r \geq 1\), so that \((f^T_t)\) is stationary.

**Assumption 2.2.** \(A_P = 1\) and \(\|A_T^r\| < 1\) for some \(r \in \mathbb{N}\).

Under Assumptions 2.1 and 2.2, \((f_t)\) becomes an \(I(1)\) process with \(m\) unit roots. In particular, (4) reduces to

\[ f^P_t = f^P_{t-1} + \varepsilon^P_t \]  \hspace{1cm} (6)
for $t = 1, 2, \ldots$, and (5) 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 (3) 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 (6) and (5) 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

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

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.

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

$$H = H_N \oplus H_S,$$  

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 (7) is orthogonal. We define $\Pi_N$ and $\Pi_S$ to be the orthogonal projections on the nonstationary and stationary subspaces $H_N$ and $H_S$ of $H$, and let

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

(8)
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.

for $t = 1, 2, \ldots$. Our subsequent theoretical development will rely on the decompositions in (7) and (8). 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 $\Pi_N$ and $\Pi_S$, and the dashed lines represent the projections $\Pi_P$ and $\Pi_T$.

2.3 Error Correction Representation

In the presence of unit roots, it is natural to consider the error correction representation of our functional autoregressive model in (1). However, the error correction model (ECM) $\Delta f_t = (A - 1)f_{t-1} + \varepsilon_t$ defined for our functional autoregressive model is not generally useful in practice since, under Assumption 2.1, its error correction operator $A - 1$ becomes non-compact. Consequently, the functional ECM for a functional autoregressive process $(f_t)$ is practically useful only when it has either infinite dimensional unit roots or it is finite-dimensional itself. The reader is referred to Chang et al. (2016a) for more details.

Here we assume that $(f_t)$ is finite dimensional and illustrate how our decomposition is related to its ECM. We let

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

and write

$$\Delta f_t = \alpha \beta' f_{t-1} + \varepsilon_t,$$  \hspace{1cm} (9)
where $\alpha$ and $\beta$, which are identified only up to their ranges, are $p \times q$, $p > q$, matrices of parameters such that the $q \times q$ matrix $\alpha'\beta$ is nonsingular. Under Assumptions 2.1 and 2.2, we may write the FAR in (1) as the ECM in (9), where $(f_t)$ is I(1), and $(\beta'f_t)$ is stationary with each column of $\beta$ representing a cointegrating relationship in $(f_t)$. In our subsequent discussion, we denote by $\alpha_\perp$ and $\beta_\perp$ the $p \times (p-q)$ matrices such that $\alpha_\perp \alpha_\perp' = 0$ and $\beta_\perp \beta_\perp' = 0$, where $\alpha_\perp$ and $\beta_\perp$ are again identified only up to their ranges. In this example, we have $H = \mathbb{R}^p$.

For $(f_t)$ generated by the ECM in (9), 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 $H = \mathbb{R}^p = \mathcal{R}(\alpha) \oplus \mathcal{R}(\beta_\perp)$. Furthermore, it follows that

$$H_P^* = H_T^1 = \mathcal{R}(\alpha_\perp) \quad \text{and} \quad H_T^* = H_P^1 = \mathcal{R}(\beta),$$

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 $\Pi_P$ and $\Pi_T$ 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 $\Pi_N$ and $\Pi_S$ 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 and Prediction

In this section, we show how we may estimate the functional autoregressive model with unit roots and make forecast with this model. In addition, we develop the asymptotic theories for the autoregressive operator estimator as well as for the predictor. Our method relies on the functional principal component analysis for the process $(f_t)$. Suppose that we observe $(f_t)$ for $t = 1, \ldots, T$. We define the unnormalized sample variance operator $Q$ and
first order autocovariance operator \( P \) by

\[
Q = \sum_{t=1}^{T} (f_{t-1} \otimes f_{t-1})
\]  

(10)

and

\[
P = \sum_{t=1}^{T} (f_t \otimes f_{t-1}).
\]  

(11)

Since \( Q \) is self-adjoint, it has real eigenvalues \( \hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \), with the corresponding eigenvectors \( \hat{v}_1, \hat{v}_2, \ldots \). Note that, since \( Q \) is positive semi-definite, \( \hat{\lambda}_k \geq 0 \) for all \( k = 1, 2, \ldots \). We normalize \( \hat{v}_k \) so that \( \|\hat{v}_k\| = 1 \) for \( k = 1, 2, \ldots \), and represent \( Q \) as

\[
Q = \sum_{k=1}^{T} \hat{\lambda}_k (\hat{v}_k \otimes \hat{v}_k)
\]

in its spectral form.

3.1 Functional Principal Component Analysis

In this section we study the properties of the estimated eigenvalues \( \hat{\lambda}_k \). We write

\[
Q = T^2 Q_{NN} + T Q_{NS} + T Q_{SN} + T Q_{SS},
\]  

(12)

where

\[
Q_{NN} = \frac{1}{T^2} \sum_{t=1}^{T} (f_{t-1}^N \otimes f_{t-1}^N),
\]

\[
Q_{SS} = \frac{1}{T} \sum_{t=1}^{T} (f_{t-1}^S \otimes f_{t-1}^S),
\]

\[
Q_{NS} = \frac{1}{T} \sum_{t=1}^{T} (f_{t-1}^N \otimes f_{t-1}^S),
\]

and \( Q_{SN} = Q_{NS}^* \). We first introduce a few concepts from Kuelbs (1973).

A measure \( \mu \) on a real separable Banach space \( B \) is called a Gaussian measure if for every continuous linear functional \( v \) on \( B \) and for each measurable set in \( \mathcal{B}(B) \), the measure \( \mu v^{-1} \) defined by \( \mu v^{-1}(E) = \mu(v^{-1}(E)) \) is a mean zero Gaussian measure on \( \mathbb{R} \) with variance \( \int_B v^2 d\mu \). Specifically, an \( H \)-valued random element \( f \) is said to be Gaussian with mean \( m_f \) and variance operator \( S \) if its induced probability measure on \( H \) is Gaussian and for all
the random variable \( \langle v, f - m_f \rangle \) has variance \( \langle v, Sv \rangle \). We denote such random elements by \( N(m_f, S) \). For a Gaussian measure \( \mu \) on \( B \), define a family \( \{ \mu_t : t \geq 0 \} \) of Gaussian measures on \( B \) by

\[
\mu_t(E) = \begin{cases} 
\delta_0(E) & t = 0, \\
\mu(E/\sqrt{t}) & t > 0.
\end{cases}
\]

Let \( C_B \) be the set of all continuous functions \( \omega : [0,1] \to B \) such that \( \omega(0) = 0 \). Let \( C_B \) be the \( \sigma \)-algebra on \( C_B \) generated by the mappings \( \omega \mapsto \omega(t), t \in [0,1] \). Then there is a unique probability measure \( P \) on \( C_B \) such that for any \( 0 = t_0 < t_1 < \cdots < t_n \leq 1 \), the \( B \)-valued random variables \( \omega(t_j) - \omega(t_{j-1}), j = 1, \ldots, n \) are independent and \( \omega(t_j) - \omega(t_{j-1}) \) has distribution \( \mu_{t_j-t_{j-1}} \) on \( B \). We call the stochastic process \( \{ W_t : t \geq 0 \} \) defined on \( (C_B, C_B, P) \) by \( W(t, \omega) = \omega(t) \) a \( \mu \)-Brownian motion on \( B \). By definition, \( W_t \) has stationary independent mean zero Gaussian increments. Specifically, if \( W_t \) is Brownian motion on \( H \), \( \langle v, W_t \rangle \) is a Brownian motion on \( \mathbb{R} \).

The following lemma gives an invariance principle in \( H \).

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

\[
W_T(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \varepsilon_t
\]

for \( r \in [0,1] \), then \( W_T \to_d W \) as \( T \to \infty \), where \( W \) is Brownian motion on \( H \) with variance operator \( \Sigma \).

We are now ready to establish the asymptotics for \( Q \), which is the unnormalized sample variance operator for \( (f_t) \), \( f_t = f_t^N + f_t^S \). 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 \( (\varepsilon_t^P) \), 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

\[
W_{PT}(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \varepsilon_t^P
\]

for \( r \in [0,1] \), then \( W_{PT} \to_d W_P \) as \( T \to \infty \), where \( W_P \) is a Brownian motion on \( H_P \) with variance \( \Sigma_P = \Pi_P \Sigma \Pi_P \), which is finite dimensional. Therefore, if properly normalized, \( (f_t^P) \) behaves like \( W_P \) in the limit. 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 
\[
\sum_{i=0}^{\infty} A_i^T \Sigma_T A_i^T
\] 
where \(\Sigma_T = \Pi_T \Sigma \Pi_T^*\). Since 
\[f^S_t = \Pi_S f^T_t,\]
the variance operator of \((f^S_t)\) is given by \(\Pi_S (\sum_{i=0}^{\infty} A_i^T \Sigma_T A_i^T) \Pi_S\).

The following Lemma 3.2, Lemma 3.3 and Theorem 3.4 are the analogues of Lemma 3.1, Lemma 3.2 and Theorem 3.3 in Chang et al. (2016d) respectively.

Lemma 3.2. Let Assumptions 2.1 and 2.2 hold. Then 
\[Q_{NN} \rightarrow_d \int_0^1 (W_P \otimes W_P)(r)dr\]
and 
\[Q_{SS} \rightarrow_p \Pi_S \left( \sum_{i=0}^{\infty} A_i^T \Sigma_T A_i^T \right) \Pi_S\]
as \(T \rightarrow \infty\). Moreover, 
\[Q_{NS} = Q_{SN}^* = O_p(1)\]
for large \(T\).

Suppose at this moment that the dimension \(m\) of the subspace \(H_N\) is known. We may estimate the nonstationary subspace \(H_N\) by 
\[\hat{H}_N = \bigvee_{i=1}^{m} \hat{v}_i,\]
and the stationary subspace \(H_S\) by \(\hat{H}_S = \hat{H}_N^\perp\), the orthogonal complement of \(\hat{H}_N\). We denote the orthogonal projections on \(\hat{H}_N\) and \(\hat{H}_S\) by \(\hat{\Pi}_N\) and \(\hat{\Pi}_S\), respectively. Note that \(\hat{\Pi}_N + \hat{\Pi}_S = 1\).

Lemma 3.3. Let Assumptions 2.1 and 2.2 hold. Then 
\[\hat{\Pi}_N = \Pi_N + O_p(T^{-1})\]
and 
\[\hat{\Pi}_S = \Pi_S + O_p(T^{-1})\]
for large \(T\).

For notation simplicity, we define 
\[\Sigma_{NN} = \int_0^1 (W_P \otimes W_P)(r)dr\]
and 
\[\Sigma_{SS} = \Pi_S \left( \sum_{i=0}^{\infty} A_i^T \Sigma_T A_i^T \right) \Pi_S,\]
so that \( Q_{NN} \to_d \Sigma_{NN} \) and \( Q_{SS} \to_p \Sigma_{SS} \). Note that \( \Sigma_{NN} \) is a random operator on \( H_N \), while \( \Sigma_{SS} \) is a deterministic operator on \( H_S \). Let \((\lambda_i(\Sigma_{NN}), v_i(\Sigma_{NN}))\) be the pairs of eigenvalues and eigenvectors of \( \Sigma_{NN} \) such that \( \lambda_i(\Sigma_{NN}) \)'s are in descending order. Similarly let \((\lambda_i(\Sigma_{SS}), v_i(\Sigma_{SS}))\) be the ordered pairs of eigenvalues and eigenvectors of \( \Sigma_{SS} \) such that \( \lambda_i(\Sigma_{SS}) \)'s are in descending order. For convenience, let \((\lambda_i, v_i) = (\lambda_i(\Sigma_{NN}), v_i(\Sigma_{NN}))\) for \( i = 1, 2, \ldots, m \), and \((\lambda_{i}, v_{i}) = (\lambda_{i-m}(\Sigma_{SS}), v_{i-m}(\Sigma_{SS}))\) for \( i = m+1, m+2, \ldots \). We assume that the eigenvalues \( (\lambda_i)_{i>m} \) are different from each other.\(^2\) The eigenvectors are normalized so that \( \|v_k\| = 1 \) and that \( \langle \hat{v}_k, v_k \rangle \geq 0 \) for definiteness as in Bosq (2000).

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

\[
(T^{-2}\hat{\lambda}_i, \hat{v}_i) \to_d (\lambda_i(\Sigma_{NN}), v_i(\Sigma_{NN}))
\]

as \( T \to \infty \) jointly for \( i = 1, 2, \ldots, m \), and

\[
(T^{-1}\hat{\lambda}_{m+i}, \hat{v}_{m+i}) \to_p (\lambda_i(\Sigma_{SS}), v_i(\Sigma_{SS}))
\]

as \( T \to \infty \) for \( i = 1, 2, \ldots \).

Chang et al. (2016d) develop a consistent test for the number of unit roots \( m \) in the functional process, i.e., the dimension of the nonstationary subspace. Their method, as suggested in the above theorem, utilizes the fact that the eigenvalues associated with the two subspaces diverge at different rates. Therefore, we may use their method to determine the dimension \( m \) of the nonstationary subspace. The estimation of \( m \) is essential only in the sense that one needs to know an upper bound of \( m \) so that our choice of \( K \), the number of principal components preserved when we invert \( Q \), does not fall below \( m \) (see below). Of course, we may also use other methods to estimate the dimension of the nonstationary subspace. The exact value of \( m \) does not affect the asymptotic properties of the autoregressive operator estimator as long as we guarantee that \( K \) is larger than \( m \). In practice, we may always choose \( K \) large enough so that \( K > m \) with probability arbitrarily close to one.

### 3.2 Autoregressive and Error Variance Operators

Let \( K_T \) be a positive integer valued function of \( T \) such that \( m < K_T < T \) and that \( K_T \to \infty \) as \( T \to \infty \). For notation simplicity, we write \( K \) instead of \( K_T \) in what follows.

---

\(^2\)We could potentially allow for multiplicity. However in that case the eigenvectors could not be uniquely identified even after normalization, which introduces technical complications.
However, readers should keep in mind the dependence of $K$ on $T$. Now we define

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

and define our FAR estimator of $A$ as

$$\hat{A} = PQ^+.$$ (14)

Note that in the definition of $\hat{A}$, instead of using $Q^{-1}$, we use $Q^+$, the inverse of $Q$ restricted on the subspace $\hat{H}_K$ of $H$ spanned by $\hat{v}_1, \ldots, \hat{v}_K$. This is because in the infinite dimensional setting $Q^{-1}$ becomes extremely non-smooth as $T \to \infty$, which implies that the inference based on $Q^{-1}$ is highly unstable. This is the so called ill-posed inverse problem. To regularize $Q^{-1}$, we need to choose $K$ so that $Q^+$ is much smoother than $Q^{-1}$, and at the same time not too “far away” from $Q^{-1}$.

To establish the consistency of the autoregressive operator estimator, we first introduce the following lemma from Hu et al. (2016). We define a sequence $(\tau_k)$ for $k = m+1, m+2, \ldots$ by $\tau_{m+1} = 2\sqrt{2}(\lambda_{m+1} - \lambda_{m+2})^{-1}$ and $\tau_k = 2\sqrt{2} \max\{(\lambda_{k-1} - \lambda_k)^{-1}, (\lambda_k - \lambda_{k+1})^{-1}\}$ for $k > m+1$.

**Lemma 3.5.** Let Assumptions 2.1 and 2.2 hold. Then

$$\|Q_{SS} - \Sigma_{SS}\| = O\left(T^{-1/2} \log^{1/2} T\right) \quad a.s..$$

Moreover,

$$\sup_{k \geq m+1} T^{-1} \hat{\lambda}_k - \lambda_k \leq \|Q_{SS} - \Sigma_{SS}\|$$

and

$$\|\hat{v}_k - v_k\| \leq \tau_k \|Q_{SS} - \Sigma_{SS}\|,$$

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

The following theorem, giving the orders of the interaction terms, is useful in the proof of asymptotic properties of the autoregressive operator estimator.

**Lemma 3.6.** Let Assumptions 2.1 and 2.2 hold. Then

$$\left\| \sum_{t=1}^{T} (\varepsilon_t \otimes f^S_{t-1}) \right\| = O_p(\sqrt{T})$$
and
\[ \left\| \sum_{t=1}^{T} (\varepsilon_t \otimes f_{t-1}^N) \right\| = O_p(T) \]
for large $T$.

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).

**Assumption 3.1.** $\log T \left( \sum_{k=m+1}^{K} \tau_k \right)^2 / \left( T \lambda_K^2 \right) \to 0$ as $T \to \infty$.

Note that Assumption 3.1 does not put any actual restrictions on the time series $(f_t)$ itself. Since $\sum_{k=m+1}^{K} \tau_k$ is increasing in $K$ and $\lambda_K$ is decreasing in $K$, it merely controls how fast $K$ may grow as $T \to \infty$. That is, it only imposes a restriction on how we may choose $K$ as a function of $T$. The following theorem establishes the consistency of our autoregressive operator estimator.

**Theorem 3.7.** Let Assumptions 2.1, 2.2 and 3.1 hold. Then
\[ \left\| \hat{A} - A \right\| \to_p 0 \]
as $T \to \infty$.

Our estimator $\hat{A}$ of the autoregressive operator $A$ in (14) is therefore consistent.

Once we obtain the autoregressive operator estimator, we may obtain the residuals by
\[ \hat{\varepsilon}_t = f_t - \hat{A} f_{t-1} \]
and estimate $\Sigma$ by
\[ \hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} \left( \hat{\varepsilon}_t \otimes \hat{\varepsilon}_t \right) \]

The following corollary is an obvious consequence of Theorem 3.7.

**Corollary 3.8.** Let Assumptions 2.1, 2.2 and 3.1 hold. Then
\[ \left\| \hat{\Sigma} - \Sigma \right\| \to_p 0 \]
as $T \to \infty$.

The following two theorems establish the convergence rate of our estimator $\hat{A}$. It turns
out that $\hat{A}$ converges at different rates on different subspaces. We let

$$\hat{\Pi}_K = \sum_{k=1}^{K} \hat{v}_k \otimes \hat{v}_k,$$

i.e., the orthogonal projection on the subspace $\hat{H}_K$ of $H$ spanned by $\hat{v}_1, \ldots, \hat{v}_K$.

**Theorem 3.9.** Let Assumptions 2.1, 2.2 and 3.1 hold. Then

$$T \left( \hat{A} - A\hat{\Pi}_K \right) \Pi_N \to_d \left( \int_0^1 (dW \otimes W_P)(r) \right) \left( \int_0^1 (W_P \otimes W_P)(r)dr \right)^{-1}$$

as $T \to \infty$.

Theorem 3.9 provides the limit distribution of $\hat{A} - A\hat{\Pi}_K$ on the nonstationary subspace $H_N$, which is given as a function of Brownian motions. As expected, if $H$ is one-dimensional, the limit distribution reduces to the Dickey-Fuller distribution. Unfortunately, as observed in Mas (2007), it is impossible for $\hat{A} - A\hat{\Pi}_K$ to weakly converge in operator norm on the stationary subspace $H_S$. However, we may still establish the pointwise asymptotic normality of $\hat{A} - A\hat{\Pi}_K$. This is shown below.

**Theorem 3.10.** Let Assumptions 2.1, 2.2 and 3.1 hold. Then, for any $v \notin H_N$,

$$\sqrt{\frac{T}{s_K(v)}} \left( \hat{A} - A\hat{\Pi}_K \right) \Pi_S v \to_d N(0, \Sigma)$$

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

Note that the convergence rate of $(\hat{A} - A\hat{\Pi}_K)\Pi_S v$ for $v \notin H_N$ depends on $v$. Specifically, $(\hat{A} - A\hat{\Pi}_K)\Pi_S v$ converges at the parametric $\sqrt{T}$-rate if $\sum_{k=1}^{\infty} \lambda_k^{-1} \langle v_k, v \rangle^2 < \infty$, and converges at a rate slower than $\sqrt{T}$ if $\sum_{k=1}^{\infty} \lambda_k^{-1} \langle v_k, v \rangle^2 = \infty$.

Theorems 3.9 and 3.10 show that our estimator $\hat{A}$ 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) $\lambda_k$ is convex in $k$ for $k$ large enough,

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

(c) $\sum_{i=m+1}^{K} \sum_{j=m+K+1}^{\infty} \lambda_i \lambda_j / (\lambda_i - \lambda_j)^2 = o(K)$
The conditions in Assumption 3.2 are not very restrictive. The condition in (a) is very mild and is satisfied by many sequences of eigenvalues decaying at polynomial and exponential rates. The condition in (b) holds as long as $K = O(T^{1/5-\delta})$ for any $\delta > 0$, and $K$ does not grow too fast as $T \to \infty$. The condition in (c) is more stringent, though not prohibitively so. For many practical applications, it appears that $(\lambda_k)$ decays geometrically and we may set $\lambda_k = \rho^k$ for some $0 < \rho < 1$. In this case, we may easily deduce that

$$\sum_{i=m+1}^{K} \sum_{j=m+K+1}^{\infty} \lambda_i \lambda_j / (\lambda_i - \lambda_j)^2 = O(1).$$

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

$$(A\hat{\Pi}_K - A)\Pi_N = o_p(T^{-1/2}K^{1/2}),$$

and, for any $v \notin H_N$,

$$(A\hat{\Pi}_K - A)\Pi_S v = o_p(T^{-1/2}K^{1/2}) + O(\|(1 - \Pi_K)v\|)$$

for large $T$.

Theorem 3.11 provides the orders of the bias terms in our autoregressive operator estimator $\hat{A}$ on the nonstationarity subspace $H_N$ and the stationarity subspace $H_S$. On both $H_N$ and $H_S$, the bias terms become negligible as long as $K \to \infty$ as $T \to \infty$. Note that $(1 - \Pi_K)v \to 0$ as $K \to \infty$ for any $v \in H$.

### 3.3 Forecast

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

$$\hat{f}_{T+1} = \hat{A} f_T,$$

where $\hat{A}$ is the estimated autoregressive operator. 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 $\hat{A}$ using data only up to time $T - 1$ to avoid technicalities.

**Assumption 3.3.** We assume that

(a) $\|\Sigma_{SS}^{-1/2} \hat{A}\| < \infty$, and

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

19
Loosely put, condition (a) requires that $A$ be at least as smooth as $\Sigma_{SS}^{1/2}$ 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 $M = 3$.

**Lemma 3.12.** Let Assumptions 2.1, 2.2, 3.1 and 3.3 hold. Then

$$\sqrt{T/K} \left( \hat{A} f_T - A \hat{\Pi}_K f_T \right) \to_d N(0, \Sigma)$$

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

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

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

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

**Theorem 3.13.** Let Assumptions 2.1, 2.2, 3.1, 3.2, 3.3 and 3.4 hold. Then

$$\sqrt{T/K} \left( \hat{A} f_T - A f_T \right) \to_d N(0, \Sigma)$$

as $T \to \infty$.

From Theorem 3.13 we may easily deduce that for any Gaussian $(\varepsilon_t)$, we have that

$$\hat{f}_{T+1} - f_{T+1} = \hat{A} f_T - A f_T - \varepsilon_{T+1} \approx_d N \left( 0, \left( 1 + \frac{K}{T} \Sigma \right) \right).$$

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

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

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

### 3.4 Beveridge-Nelson Decomposition

The Beveridge-Nelson decomposition can also be estimated consistently. To explain how to estimate it, we assume that the number $m$ of unit roots is known. This causes no loss of
generality, since it can be estimated consistently as shown in Chang et al. (2016d). We let 
\( \hat{\alpha}_1, \ldots, \hat{\alpha}_K \) be the \( K \) nonzero eigenvalues of \( \hat{A} \) such that \( |\hat{\alpha}_1| \geq \cdots \geq |\hat{\alpha}_K| \), and define

\[
\pi_P(A) = (A - \hat{\alpha}_1)(A - \hat{\alpha}_2) \cdots (A - \hat{\alpha}_m),
\]

\[
\pi_T(A) = (A - \hat{\alpha}_{m+1})(A - \hat{\alpha}_{m+2}) \cdots (A - \hat{\alpha}_K).
\]

Then we may consistently estimate \( H_P \) and \( H_T \) by

\[
\hat{H}_P = \ker \left( \pi_P(\hat{A}) \right) \quad \text{and} \quad \hat{H}_T = \ker \left( \pi_T(\hat{A}) \right)
\]

respectively, where \( \ker (\cdot) \) denotes the kernel or null space of an operator.

If we do not have repeated eigenvalues of \( \hat{A} \), i.e., \( \hat{\alpha}_1 \neq \cdots \neq \hat{\alpha}_K \), which will always be the case in applications with real data, \( \hat{H}_P \) and \( \hat{H}_T \) are given simply as

\[
\hat{H}_P = \bigvee_{k=1}^m \hat{u}_k \quad \text{and} \quad \hat{H}_T = \bigvee_{k=m+1}^K \hat{u}_k,
\]

where \( \hat{u}_1, \ldots, \hat{u}_K \) are the eigenvectors associated with eigenvalues \( \hat{\alpha}_1, \ldots, \hat{\alpha}_K \). Note that

\[
\hat{H}_P \subset \hat{H}_K \quad \text{and} \quad \hat{H}_T \subset \hat{H}_K,
\]

if \( |\hat{\alpha}_k| \neq 0 \) for all \( k = 1, \ldots, K \). The following theorem gives the consistency of the two estimated subspaces.

**Theorem 3.14.** Let Assumptions 2.1, 2.2 and 3.1 hold. Then

\[
\left\| \hat{\Pi}_P - \Pi_P \right\| = o_p(1) \quad \text{and} \quad \left\| \hat{\Pi}_T - \Pi_T \right\| = o_p(1).
\]

for large \( T \).

Due to Theorem 3.14, we may consistently estimate the Beveridge-Nelson decomposition of autoregressive functional time series.

Since \( H_P = H_N \), we may also estimate \( H_P \) by a consistent estimator of \( H_N \), which is defined as

\[
\hat{H}_P = \hat{H}_N = \bigvee_{k=1}^m \hat{v}_k,
\]

where \( \hat{v}_1, \ldots, \hat{v}_K \) are the eigenvectors associated with eigenvalues \( \hat{\lambda}_1, \ldots, \hat{\lambda}_K \) introduced in (13). Finally, it is also possible to consistently estimate \( H_P \) with unit roots imposed, in
which case we have

\[ \tilde{H}_P = \ker \left( (\hat{A} - 1)^m \right), \]

in place of \( \hat{H}_P \) or \( \tilde{H}_P \) introduced above.

4 Extensions

4.1 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

\[ f_t = \nu + Af_{t-1} + \varepsilon_t \]

in place of (1), where \( \nu \in H_T \) and \( A \) satisfies Assumptions 2.1 and 2.2. Note that \( \nu \) 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 \( (f^P_t) \) of \( (f_t) \) would generate a linear time trend in \( (f_t) \). We may rewrite the functional autoregressive model in (16) as

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

where \( \mu = (1 - A_T)^{-1}\nu \) on \( H_T \) and \( \mu = Ef^T_t \).

To estimate the autoregressive operator \( A \) in (16), we need to first demean the time series \( (f_t) \), where the sample mean of the time series is given by

\[ \bar{f} = \frac{1}{T} \sum_{t=1}^{T} f_t. \]

Subsequently, we denote the demeaned time series of \( (f_t) \) by \( f^\mu_t = f_t - \bar{f} \), and redefine the operators \( Q \) and \( P \) by

\[ Q = \sum_{t=1}^{T} (f^\mu_{t-1} \otimes f^\mu_t) \]

and

\[ P = \sum_{t=1}^{T} (f^\mu_t \otimes f^\mu_{t-1}), \]

and \( \hat{\lambda}_k \) and \( \hat{\phi}_k \) as the ordered eigenvalues and eigenvectors of \( Q \). Furthermore, we also
redefine the autoregressive operator estimator as in (14) where $Q^+$ is defined as in (13).

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

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

for large $T$, and consequently, on $H_T$, all our sample statistics redefined by $(f_t^\mu)$ yield the same asymptotics as those defined for $(f_t)$. Therefore, the use of $(f_t^\mu)$ just gets rid of the non-zero mean in $(f_t)$ without affecting any asymptotics on $H_T$. On the contrary, however, the effect of demeaning in $(f_t^\mu)$ continue to remain in asymptotics on $H_P$. More precisely, our previous asymptotics involving functions of $W_P$ are now represented by the same 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. (2016d) for more details.

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

**Lemma 4.1.** Let Assumptions 2.1 and 2.2 hold. Then

$$Q_{NN} \overset{d}{\rightarrow} \int_0^1 (W_P^\mu \otimes W_P^\mu)(r) dr$$

$$Q_{SS} \overset{p}{\rightarrow} \Pi_S \left( \sum_{i=0}^{\infty} A_S^i \Sigma T A_S^{*i} \right) \Pi_S$$

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

$$Q_{NS} = Q_{SN}^* = O_p(1)$$

for large $T$.

Lemma 3.3 and Lemma 3.5 continue to hold for the functional autoregression (16). Moreover, Theorem 3.4 holds with $\Sigma_{NN}$ redefined as $\Sigma_{NN} = \int_0^1 (W_P^\mu \otimes W_P^\mu)(r) dr$, and Lemma 3.6 holds with $(f_t)$ replaced by the $(f_t^\mu)$. 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

$$\hat{f}_{T+1} = \bar{f} + \hat{A} f_T^\mu$$

to reflect the required demeaning procedure.
**Theorem 4.2.** Let the assumptions in Theorem 3.7 hold. Then

\[ \| \hat{A} - A \| \to_p 0 \]

as \( T \to \infty \). If in addition the assumptions in Theorem 3.13 hold, then

\[ \sqrt{T/K} \left( (\hat{f}_{T+1} - \mu) - A(f_T - \mu) \right) \to_d N(0, \Sigma) \]

as \( T \to \infty \).

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 autoregressions continue to hold essentially without any additional assumptions.

**4.2 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 \((\hat{f}_t^\mu)\),

\[ \hat{f}_t^\mu = \hat{f}_t - \frac{1}{T} \sum_{t=1}^{T} \hat{f}_t, \]

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

We denote the estimation error 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 \((\Delta_t)\). We therefore introduce the following assumption.

**Assumption 4.1.** \(\sup_{t \geq 1} \| \Delta_t \| = O_p(1/\sqrt{T})\).

Under Assumption 4.1, \(\| \Delta_t \|\) 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, has therefore no bearing on our asymptotics. This is well expected from Chang et al. (2016d). 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 \(T\), which appears to be the case for most practical applications.
4.3 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

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

where $A_1, A_2, \cdots, A_p$ are compact operators on $H$ and $(\varepsilon_t)$ is a functional white noise that satisfies Part (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 (17) as an $H^p$-valued FAR(1) process given by

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

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

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

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

**Assumption 4.2.** $A(1)$ is not invertible, and if $A(z)$ is not invertible, then $z = 1$ or $|z| < 1$.

Define

$$M(z) = \begin{bmatrix}
0 & -1 & 0 & \cdots & 0 & 0 \\
0 & 0 & -1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & -1 \\
A_0(z) & A_1(z) & A_2(z) & \cdots & A_{p-2}(z) & A_{p-1}(z)
\end{bmatrix}.$$
where \( A_0(z) = 1 \) and \( A_i(z) = zA_{i-1}(z) - A_i \) for \( i \geq 0 \), and define

\[
N(z) = \begin{bmatrix}
  1 & z & z^2 & \ldots & z^{p-2} & z^{p-1} \\
  0 & 1 & z & \ldots & z^{p-3} & z^{p-2} \\
  \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
  0 & 0 & 0 & \ldots & 1 & z \\
  0 & 0 & 0 & \ldots & 0 & 1 \\
\end{bmatrix}.
\]

Then it follows that

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

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 4.2 implies that \( \sup \lambda(A) \setminus \{1\} < 1 \). Furthermore, since \( \sup \lambda(A) \setminus \{1\} = \lim_{r \to \infty} \| A_r^T \|^1/r \), there exists \( r \in \mathbb{N} \) such that \( \| A_r^T \| < 1 \). Consequently, Assumption 2.2 holds for the model (18). 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.

## 5 Term Structure of US Government Bond Yields

The term structure of interest rates is one of the most important topics in finance and macroeconomics. In finance, it is the foundation of asset pricing. In macroeconomics, it plays a central role in agents’ intertemporal choices and risk sharing. 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 bond at time \( t \) that promises to pay \$1 \( \tau \) years ahead. The price \( P_t(\tau) \) can be obtained through compound discounting with the yield to maturity \( y_t(\tau) \):

\[
P_t(\tau) = \exp(-\tau y_t(\tau)).
\]

The ratio of change in the bond’s price at any future time \( t + \tau \) defined by

\[
f_t(\tau) = \frac{P_t'(\tau)}{P_t(\tau)}
\]
is called the (instantaneous) forward rate, which gives the instantaneous interest rate \( \tau \)-period ahead of \( t \). The graph of \( f_t \) as a function of the maturity \( \tau \) is called the forward rate curve. With \( f_t \), we are able to obtain the present value or the expected future value of any cash flow. Therefore, the forward rate curve is widely used in asset pricing. In this section, we study the dynamics of the forward rate curves of US government bonds.

Unfortunately, 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

\[
f_t(\tau) = \beta_0 t + \beta_1 e^{-\tau/\gamma_1 t} + \beta_2 \frac{\tau}{\gamma_1 t} e^{-\tau/\gamma_1 t} + \beta_3 \frac{\tau^2}{\gamma_2 t} e^{-\tau/\gamma_2 t},
\]

where \( \beta_0, \beta_1, \beta_2, \gamma_1, \gamma_2 \) are the parameters to be estimated in each period. They estimate the forward rate curves at daily frequency from 1961 on.\(^3\) We use their estimated end-of-month forward rate curves from January 1985 to September 2016. Figure 3 plots the time series of the forward rate curves.

Figure 3: Time Series of Forward Rate Curves

![Figure 3: Time Series of Forward Rate Curves](image)

\textit{Notes:} This figure plots the time series of the end-of-month US government bond forward rate curves from January 1985 to September 2016. Data source: Gürkaynak et al. (2007).

Two features in the time series of forward rate curves are worth mentioning. First, 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

---

\(^3\)The Federal Reserve Board maintains a web page which posts the update of the estimated forward rate curves quarterly. See http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html.
deterministic trend, or the deterministic trend should be very weak. In this paper, we assume that the trend is stochastic, and use (1) to model the demeaned forward rate curve process. The second feature to note is that the plot clearly indicates the existence of the zero-lower-bound for short term interest rate after the 2007 financial crisis.\footnote{It could be possible that there was a structural change in the bond market due to changes in the macro economy after the 2007 financial crisis. To formally investigate this possibility, we need a test for structural changes for functional time series. This exceeds the scope of our paper. In the following analysis, we assume that there was no structural change within the time span considered.}

We first demean the data by subtracting the sample mean from the time series of forward rate curves. We then represent the 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 representations of the operators $Q$ and $P$ defined in (10) and (11) as two 1037-by-1037 matrices and then obtain the eigenvalues and eigenvectors of $Q$. We estimate $A$ as in (14). The test developed in Chang et al. (2016d) suggests two unit roots in the process. Therefore we set $\hat{m} = 2$. In addition, we set $K = 5$ to obtain the best rolling out-of-sample forecast performance, in which we use the last one-fifth of periods as the prediction periods. Figure 4 plots the ten largest eigenvalues of the unnormalized sample variance operator $Q$ in descending order. It can be seen that the eigenvalues decay to zero very fast. The first five principal components explain 99.9\% of variations in the data, which justifies our choice of the value of $K$.

Figure 4: Scree Plot for the Forward Rate Curves

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{scree_plot.png}
\caption{Scree Plot for the Forward Rate Curves}
\end{figure}

Notes: This figure plots in descending order the ten largest eigenvalues of the unnormalized sample variance operator $Q$.

A class of interesting variables that can be written as linear functionals of the forward rate curve is the yield to maturity $y_t(\tau)$. It measures the average interest rate at time $t$ of holding a zero coupon bond until $\tau$ periods ahead. $y_t(\tau)$ may be generated by $v_{\tau}(x) = \frac{1}{\tau} 1\{x \leq \tau\}$ through $y_t(\tau) = \langle v_{\tau}, f_t \rangle$. We take $y_t(0.25)$, the 3-month yield to maturity, as the short term interest rate at time $t$. Similarly, we take $y_t(3)$ and $y_t(10)$ as the medium...
term interest rate and the long term interest rate at time \( t \) respectively.

Figure 5: Response Functions of Interest Rate of Different Terms to Changes in the Past Forward Rate

![Graph showing response functions of interest rate](image)

Notes: This figure plots the response functions of the short, medium, and long term interest rates one month ahead to shocks to the current forward rate.

Figure 5 plots how the short, medium and long term interest rates one month ahead respond to Dirac-\( \delta \) impulses to the current instantaneous forward rates for different maturities. The short term interest rate responds mainly to changes in the last month’s forward rates for maturities between zero and two years, i.e., the expected interest rate for the near future. The shorter the maturity, the greater the response. Shocks to the forward rates for long maturities do not affect the future short term interest rate much. The medium term interest rate one month ahead responds mainly to changes in the current forward rates for maturities between zero and four years, with the most impact coming from the changes in the forward rates for maturities around one year. The future medium term interest rate thus responds mainly to changes in the current expected interest rate for the near and medium future. In addition, it responds slightly negatively to changes in the forward rates
for maturities between four to eight years, and responds slightly positively to changes in the forward rates for maturities longer than eight years. The long term interest rate one month ahead responds to shocks to the current forward rates for almost all maturities, with the shocks to the long forward rates generating the most impact. In sum, changes in the current forward rates affect future short, medium and long term interest rates in different ways.

Figure 6: Response Functions of the Interest Rates at Different Future Dates to Changes in the Current Forward Rate

Notes: This figure plots the response functions of the one-month interest rates one months, three months, six months and one year later to shocks to the current forward rate.

An important question researchers on the term structure of interest rates seek to answer is whether the current forward rates forecast future interest rates. In this paper, we try to
answer this question by looking at the impulse responses of \( y_{t+\tau}(1/12) \) to shocks to \( f_t \) for different \( \tau \)'s. We choose a very short maturity, i.e., a month, so that the yield is close to the spot rate. If economic agents’ expectations are rational, or at least partially rational so that the current forward rates have some power in forecasting future interest rates, then \( y_{t+\tau}(1/12) \) should respond strongly to shocks to \( f_t \) for maturities around \( \tau \) years. The response functions in Figure 6 support the idea that the current forward rates predict the future interest rates: the interest rates one month, three months, six months and one year ahead respond most strongly to shocks to the current forward rate for maturities one month, three months, six months and one year, respectively.

Figure 7: Decomposition of the Time Series of the Current Forward Rate

Notes: The top left panel presents the demeaned time series of forward rate curves. The top right panel presents the permanent component of the time series. The bottom left panel presents the transitory component of the time series. The bottom right panel presents the component not captured by the top right and the bottom left panel due to truncation at the \( K \)-th principal component.

Theorem 2.3 implies that the time series generated by the model (1) can be decomposed into a random walk process and a stationary process. The two processes have the inter-
pretation as the permanent component and the transitory component of the time series respectively. In Section 3.4 we provide a method to consistently estimate the two components. Figure 7 gives the decomposition results for the time series of the forward rate curves. The top left panel presents the demeaned time series of the forward rate curves. The top right panel presents the permanent component of the time series. The bottom left panel presents transitory component of the time series. The bottom right panel presents the component not captured by the top right and the bottom left panel due to truncation at the \( K \)-th principal component. It can be easily seen that the permanent component tracks the trend in the time series of the forward rate curves very well, and the transitory component is much more stationary than the original time series, and is centered at around zero. The part that is not captured by the permanent and the transitory components is negligible. This once again justifies our choice for the value of \( K \).

In the study of the term structure of interest rates, factor models have been widely used. For example, Diebold and Li (2006) model the forward rate curves through

\[
 f_t(\tau) = \beta_1 t + \beta_2 e^{-\gamma_1 \tau} + \beta_3 t e^{-\gamma_2 \tau}
\]

and interpret the coefficients \( \beta_1 \), \( \beta_2 \) and \( \beta_3 \) respectively as the level, slope and curvature factors. Correspondingly, \( 1, e^{-\gamma_1 \tau} \) and \( \gamma_2 t e^{-\gamma_2 \tau} \) are interpreted as the level, slope and curvature factor loadings.\(^5\) Due to the singular value decomposition of compact operators, our FAR model also provides a set of factors for the yield curve. To be specific, note that we may write

\[
 f_t^P = f_{t-1}^P + \varepsilon_t^P = \left( \sum_{i=1}^{2} (v_i \otimes v_i) \right) f_t^P + \varepsilon_t^P
\]

and

\[
 f_t^T = A_T f_{t-1}^T + \varepsilon_t^T = \left( \sum_{i=1}^{\infty} \lambda_i^T (u_i^l \otimes u_i^r) \right) f_t^T + \varepsilon_t^T
\]

where \( v_i \) is defined as in Section 3, and \( (u_i^l) \) and \( (u_i^r) \) are two orthonormal systems in \( H_T \). Since the predictable parts of the permanent component and the transitory component lie in \( \sqrt{\sum_{i=1}^{2} v_i} \) and in \( \sqrt{\sum_{i=1}^{\infty} u_i^l} \) respectively, we may view \( v_i, i = 1,2 \) as the permanent factor loadings, and \( u_i^l \) as the transitory factor loadings. Correspondingly, the we may take \( \langle f_t^P, v_i \rangle \) as the permanent factors and \( \langle f_t^T, u_i^l \rangle \) as the transitory factors. We label these factors as the functional autoregression (FAR) factors.

For comparison, we also consider an alternative set of factors given by \( \langle f_t, v_i \rangle \). The

\(^5\)In their data analysis, they set \( \gamma_1 \) to be a constant for all \( t \). That is, there factor loadings are not time varying.
corresponding factor loadings are given by the $v_i$'s. Since the first $i$-th principal component of $f_t$ is given by $\langle f_t, v_i \rangle v_i$, we call such factors the functional principal component analysis (FPCA) factors. Figures 8 and 9 plot the FAR and the FPCA factors and their loadings respectively. We find that the permanent FAR factors are very close to the permanent FPCA factors, while the transitory FAR factors are different from the transitory FPCA factors. Similar results hold for the factor loadings. This implies that the FAR factors and the FPCA factors may pick up different dynamics in the transitory component of the forward rate curves.

Figure 8: Decomposition of the Time Series of the Current Forward Rate

![Factor 1](image1)
![Factor 2](image2)
![Factor 3](image3)
![Factor 4](image4)

Notes: The four panels present the first four FAR and FPCA factors. Factors 1 and 2 are the permanent factors. Factor 3 and 4 are the first two transitory factors.

To further investigate the economic meanings of these factors, we search for their most relevant variables from a large set of macroeconomic and financial variables, using the adaptive least absolute shrinkage and selection operator (LASSO). We consider the variables from the FRED-MD database constructed by McCracken and Ng (2016). This database, updated regularly, currently contains monthly series of 128 variables on output and income, labor markets, consumption, production and inventories, money and credit, interest rate and exchange rates, prices and stock markets from 1959:1 to 2016:8. Since there are missing values in many series, to maximize the number of variables used in our search, we restrict the time frame to 1985:1-2016:3. Within this time frame, only the series ACOGNO (new orders
Figure 9: Decomposition of the Time Series of the Current Forward Rate

Notes: The top panel presents the first four FAR factor loadings. The bottom panel presents the first four FPCA factor loadings.

for consumer goods) has missing values. We exclude this series in our analysis. We note that there is multicollinearity among the variables in the FRED-MD database. For example, the effective federal funds rate (FEDFUNDS), the one-year treasury constant maturity rate (GS1) and the one-year treasury constant maturity minus the federal funds rate (T1YFFM) are all included in the dataset. Zou and Hastie (2005) point out that LASSO tends to select arbitrarily one variable from a group of variables that are highly correlated. We therefore remove some variables from the FRED-MD database in our analysis, hoping to alleviate the multicollinearity issue. We end up with 82 candidate variables. See Appendix B for a complete list of variables included in the candidate set. All variables are stationarized using the transformation codes provided by the database.

The adaptive LASSO, proposed by Zou (2006), yields solutions that are sparse in the regression coefficients by adding a penalizing term to the ordinary least square estimation. If we use $y$ to denote a factor and $X$ to denote the set of candidate variables, the adaptive LASSO estimator is given by

$$
\hat{\beta}_L = \arg \min_{\beta} (y - X\beta)'(y - X\beta) + \lambda \sum_{i=1}^{N} \frac{|\hat{\beta}_i|}{|\hat{\beta}_i|}
$$
where $\lambda$ is a nonnegative regularization parameter, $N$ is the number of candidate variables, $\beta_i$ is the coefficient of the $i$-th variable and $\hat{\beta}_i$ is any consistent estimator of $\beta_i$. In this paper we set $\hat{\beta}_i$ to be the ridge regression estimator and choose $\lambda$ to minimize the Akaike information criterion (AIC). In our analysis, the first two factors are stationarized by taking the first difference.

Table 1: Variables Selected by Adaptive LASSO

<table>
<thead>
<tr>
<th>FAR Factors</th>
<th>FPCA Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor 1</strong></td>
<td>10-year treasury bill rate</td>
</tr>
<tr>
<td>10-year treasury bill rate</td>
<td>none</td>
</tr>
<tr>
<td><strong>Factor 2</strong></td>
<td>none</td>
</tr>
<tr>
<td>none</td>
<td></td>
</tr>
<tr>
<td><strong>Factor 3</strong></td>
<td>all employees: wholesale trade</td>
</tr>
<tr>
<td>new private housing permits</td>
<td></td>
</tr>
<tr>
<td>Canada/US foreign exchange rate</td>
<td></td>
</tr>
<tr>
<td><strong>Factor 4</strong></td>
<td>all employees: construction</td>
</tr>
<tr>
<td>new private housing permits</td>
<td></td>
</tr>
<tr>
<td>1-year treasury bill rate</td>
<td></td>
</tr>
<tr>
<td>volatility index</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 reports the variables selected by the adaptive LASSO for the FAR factors and the FPCA factors. Both the first FAR factor and the first FPCA factor are related to the 10-year treasury bill rate, which implies that they are the long rate factors. The second FPCA factor is related to the 3-month treasury bill rate, while the second FAR factor is not related to any of the variables we consider. Therefore, the short rate is related to a permanent FPCA factors, but not to any permanent FAR factors. The third FAR factor relates to the labor market, the housing market and the foreign exchange market, and the fourth FAR factor relates to the labor market, the housing market, the short rate and the volatility index. It is expected that forward rate curve factors relate to the short rate, the exchange rate market and the volatility index. They may relate to the housing market through mortgages. Furthermore, our factors are shown to relate to business cycles through employment. A possible channel could be that employment affects expectations for future monetary policy and therefore future interest rates. On the other hand, the two stationary FPCA factors fail to relate themselves to any of the candidate variables. Table 2 presents
the OLS estimation results for the two sets of factors, using the selected variables as the covariates. We report the coefficient estimates, their standard errors and the \( p \)-values.

<table>
<thead>
<tr>
<th>Factor 1</th>
<th>coefficient</th>
<th>s.e.</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-year treasury bill rate</td>
<td>22.43</td>
<td>1.726</td>
<td>0.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor 3</th>
<th>coefficient</th>
<th>s.e.</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>all employees: wholesale trade</td>
<td>668.8</td>
<td>206.5</td>
<td>0.001</td>
</tr>
<tr>
<td>new private housing permits</td>
<td>5.090</td>
<td>2.302</td>
<td>0.028</td>
</tr>
<tr>
<td>Canada/US foreign exchange rate</td>
<td>-73.96</td>
<td>26.12</td>
<td>0.005</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor 4</th>
<th>coefficient</th>
<th>s.e.</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>all employees: construction</td>
<td>96.92</td>
<td>47.65</td>
<td>0.043</td>
</tr>
<tr>
<td>new private housing permits</td>
<td>3.757</td>
<td>0.785</td>
<td>0.000</td>
</tr>
<tr>
<td>1-year treasury bill rate</td>
<td>6.374</td>
<td>1.123</td>
<td>0.000</td>
</tr>
<tr>
<td>volatility index</td>
<td>-0.104</td>
<td>0.033</td>
<td>0.002</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor 2</th>
<th>coefficient</th>
<th>s.e.</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-month treasury bill rate</td>
<td>11.17</td>
<td>2.335</td>
<td>0.000</td>
</tr>
</tbody>
</table>

We also consider two alternative sets of candidate variables (see Appendix B) and report the selection results using the Akaike information criterion and the Bayesian information criterion in Appendix C. Though there are differences in the selected variables, the main pattern remains. The first nonstationary FAR factor is related to the long rate. The stationary FAR factors are related to the short rate, the foreign exchange market, the housing market and the volatility index. Finally, the stationary factors relate to business cycles through employment.

## 6 Simulations

To evaluate the performance of our predictor, we conduct simulations to check prediction precision. To begin, we simulate time series of functions that follow the model (1). The simulation of data is based on the estimation results of the application in Section 5 so that
it has more practical relevance than merely a computational exercise. To be specific, we first obtain the estimated components (\( \hat{f}_P t \)) and (\( \hat{f}_T t \)) as well as the residuals (\( \hat{\epsilon}_P t \)) and (\( \hat{\epsilon}_T t \)) from the above application. Then in our simulation, \( f_P t \) and \( f_T t \) are set as the zero function, and innovations in the two corresponding subspaces are bootstrapped from the residuals (\( \hat{\epsilon}_P t \)) and (\( \hat{\epsilon}_T t \)). The FAR operator on \( H_T \) are set as the estimated FAR operator for the process (\( \hat{f}_T t \)). We simulate the time series (\( f_P t \)) and (\( f_T t \)) according to (4) and (5), and then (\( f t \)) according to (3).

For each simulated \((T + 1)\)-period time series of functions, we leave the last period as the “observed” period for prediction assessment, and use the first \( T \) periods to estimate the FAR operator as well as to make prediction. We denote the prediction for period \( T + 1 \) by \( \hat{f} \), and the actually “observed” function at period \( T + 1 \) by \( f \). To evaluate the deviation of the prediction from the actually observed, we use the mean square error (MSE) criterion.

We label our predictor as the FAR predictor. To compare our predictor with other potential predictors, we consider two benchmarks: the LAST predictor and the AVE predictor. The LAST predictor uses \( f_T \) as the prediction for \( f_{T+1} \) and the AVE predictor uses \( 1/T \sum_{t=1}^{T} f_t \) instead. They are respectively the best predictors if the true data generating process is a nonstationary martingale process and if the true data generating process is an iid process. Our simulation results show that the FAR predictor outperforms the other two predictors in the MSE. The AVE predictor performs horribly since there is strong persistence in the process. The LAST predictor also performs poorly since it fails to differentiate the stationary component from the nonstationary component and simply treats both components as nonstationary.

To further illustrate this point, we stimulate data with scaled stationary components. To be specific, in simulating the data, we rescale the FAR operator on \( H_T \) by a factor of \( c \). The smaller the \( c \) is, the farther away the stationary component is from nonstationary. Figure 10 presents the prediction results for \( c = 0.1, 0.5 \) and 1.0. The ratio of the MSE of the FAR predictor to the MSE of the LAST predictor is presented. We consider three FAR predictors. The plain FAR predictor, labeled as “FAR” in the figure, uses the estimated \( A \) to predict. The FAR with unit root and subspace restriction predictor, labeled as “FAR\_NT” in the figure, predicts the permanent and transitory components separately by imposing the restriction that the permanent component is a random walk process. The FAR with unit root restriction predictor, labeled as “FAR\_\lambda” in the figure, uses the estimated \( A \), but with its first two eigenvalues modified to one. We find the the three FAR predictors perform roughly the same, and for all three choices of \( c \), they uniformly outperform the LAST predictor. The prediction performances of the FAR predictors are better when \( c \) is small, implying that treating the stationary component as nonstationary really introduces large prediction
Figure 10: Prediction with Simulated Data

Notes: The figure plots the ratio prediction MSE of the FAR predictors to the prediction MSE of the random walk predictors for different values of $c$, a rescale parameter of which a smaller value implies a less persistent stationary component in the simulated process. We use three FAR predictors. The plain “FAR” predictor uses the estimated $A$ to predict. The “FAR$_{NT}$” predictor predicts the nonstationary and stationary component separately by imposing the restriction that the nonstationary component is a random walk process. The “FAR$_{\lambda}$” predictor uses the estimated $A$, but with its first two eigenvalues modified to exactly one, to predict.

error, and our FAR estimator can accommodate both the nonstationary and the stationary dynamics in the process.

7 Conclusions

We build an autoregressive model for time series of random functions taking values in a Hilbert space with unit roots. 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 may estimate the autoregressive operator based on functional principal component analysis, and make predictions based on the estimated operator. The estimated operator is consistent under very mild conditions with different convergence rates and limit distributions in different subspaces, and the predictor is asymptotically normal, with a convergence rate slower than
\( \sqrt{T} \). 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 is not first order Markovian but instead follows an autoregressive process of order \( p \). We apply our method to study the term structure of the US government bond yields. We find that the future short, medium and long term interest rates respond differently to changes in the current forward rates, and the current expected interest rates predict future spot rates. In addition, we propose permanent and transitory factors that relate to the long rate, the short rate, the exchange rate, the housing market, the volatility index and the labor market. Simulations show that our predictor, which accommodates both the nonstationary and stationary dynamics in the time series, outperforms its competitors.
Appendix A Mathematical Proofs

**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 $\Gamma_P$ and $\Gamma_T$ 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 $\Pi_P + \Pi_T = 1$, $\Pi_P\Pi_T = \Pi_T\Pi_P = 0$, $\Pi_P$ is the projection onto the subspace $H_P$, $H = H_P \oplus H_T$, 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,

$$\langle (A^*_P - 1)^{d-1}v, f^P_t \rangle = \langle (A^*_P - 1)^{d-1}v, A_P f^P_{t-1} \rangle + \langle (A^*_P - 1)^{d-1}v, \varepsilon^P_t \rangle$$

$$= \langle (A^*_P - 1)^{d-1}v, f^P_t \rangle + \langle (A^*_P - 1)^{d-1}v, \varepsilon^P_t \rangle,$$

which implies that $\langle (A^*_P - 1)^{d-1}v, f^P_t \rangle$ is I(1).

For $d \geq 2$, 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
\[
\left\langle \left( A_P^* - 1 \right)^d v, f_t^P \right\rangle = \left\langle \left( A_P^* - 1 \right)^d v, A_P f_t^P \right\rangle + \left\langle \left( A_P^* - 1 \right)^d v, \varepsilon_t^P \right\rangle
\]
\[
= \left\langle \left( A_P^* - 1 \right)^d v, f_t^P \right\rangle + \left\langle \left( A_P^* - 1 \right)^d v, f_t^{P-1} \right\rangle + \left\langle \left( A_P^* - 1 \right)^d v, \varepsilon_t^P \right\rangle.
\]
This shows that \( \left\langle \left( A_P^* - 1 \right)^d v, f_t^P \right\rangle \) is I(2). By the usual mathematical induction, we may now readily show that \( \left\langle v, f_t^P \right\rangle \) is I(d), and the proof is complete. \( \square \)

**Proof of Lemma 3.1.** It follows from Theorem 2.7 in Bosq (2000) that
\[
\frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon_t \to_d \mathbb{N}(0, \Sigma)
\]
where \( \mathbb{N}(0, \Sigma) \) is an \( H \)-valued Gaussian random element with variance operator \( \Sigma \). The invariance principle then follows immediately from Corollary 1 in Kuelbs (1973). \( \square \)

**Proof of Lemma 3.2.** See Lemma 3.1 in Chang et al. (2016d). \( \square \)

**Proof of Lemma 3.3.** See Lemma 3.2 in Chang et al. (2016d). \( \square \)

**Proof of Theorem 3.4.** See Theorem 3.3 in Chang et al. (2016d). \( \square \)

**Proof of Lemma 3.5.** See Theorem 2, Lemma 3 and Theorem 4 in Hu et al. (2016). \( \square \)

**Proof of Lemma 3.6.** Let \( B \) be the closed unit ball in \( H \).
\[
\left\| \sum_{t=1}^T (\varepsilon_t \otimes f_{t-1}^S) \right\| = \sup_{v_1, v_2 \in B} \left| \left\langle v_1, \left( \sum_{t=1}^T (\varepsilon_t \otimes f_{t-1}^S) \right) v_2 \right\rangle \right|
\]
\[
= \sup_{v_1, v_2 \in B} \left| \sum_{t=1}^T \langle v_1, \varepsilon_t \rangle \langle v_2, f_{t-1}^S \rangle \right|.
\]
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,
\[
\frac{1}{\sqrt{T}} \sum_{t=1}^T \langle v_1, \varepsilon_t \rangle \langle v_2, f_{t-1}^S \rangle \to_d \mathbb{N}(0, V_S(v_1, v_2))
\]
where

\[ V_S(v_1, v_2) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E(\langle v_1, \varepsilon_t \rangle^2 \langle v_2, f_{t-1}^S \rangle^2) \]

\[ = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} (E(\langle v_1, \varepsilon_t \rangle)^2)(E(\langle v_2, f_{t-1}^S \rangle)^2) \]

\[ = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \langle v_1, E(\varepsilon_t \otimes \varepsilon_t) v_1 \rangle \langle v_2, E(f_{t-1}^S \otimes f_{t-1}^S) v_2 \rangle \]

\[ \leq \|E(\varepsilon_t \otimes \varepsilon_t)\| \|E(f_{t-1}^S \otimes f_{t-1}^S)\| \]

for all \( v_1, v_2 \in B \). Since \((\varepsilon_t)\) 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 \(1/\sqrt{T} \sum_{t=1}^{T}(\varepsilon_t \otimes f_{t-1}^S)\) 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

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

Similarly, we have that

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

By Lemma 3.1 and the remarks that follows, we have

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

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

\[ \left\| \sum_{t=1}^{T}(\varepsilon_t \otimes f_{t-1}^N) \right\| = O_p(T). \]
Proof of Theorem 3.7. Write

\[ \hat{A} - A = D_1 + D_2 \]

where

\[ D_1 = \hat{A}\hat{\Pi}_N - A\hat{\Pi}_N, \]

and

\[ D_2 = \hat{A}\hat{\Pi}_S - A\hat{\Pi}_S. \]

For the first term, we have that

\[ D_1 = \left( \sum_{t=1}^{T} (f_t \otimes f_{t-1}) \right) \left( \sum_{k=1}^{K} \hat{\lambda}_k^{-1} (\hat{v}_k \otimes \hat{v}_k) \right) \hat{\Pi}_N - A\hat{\Pi}_N \]

\[ = A \left( \sum_{t=1}^{T} (f_{t-1} \otimes f_{t-1}) \right) \left( \sum_{k=1}^{m} \hat{\lambda}_k^{-1} (\hat{v}_k \otimes \hat{v}_k) \right) + \left( \sum_{t=1}^{T} (\epsilon_t \otimes f_{t-1}) \right) \left( \sum_{k=1}^{m} \hat{\lambda}_k^{-1} (\hat{v}_k \otimes \hat{v}_k) \right) - A\hat{\Pi}_N \]

\[ = \left( \sum_{t=1}^{T} (\epsilon_t \otimes f_{t-1}) \right) \left( \sum_{k=1}^{m} \hat{\lambda}_k^{-1} (\hat{v}_k \otimes \hat{v}_k) \right). \]

Since

\[ \left\| \sum_{t=1}^{T} (\epsilon_t \otimes f_{t-1}) \right\| \leq \left\| \sum_{t=1}^{T} (\epsilon_t \otimes f_{t-1}^S) \right\| + \left\| \sum_{t=1}^{T} (\epsilon_t \otimes f_{t-1}^N) \right\| = O_p(T). \]

and

\[ \hat{\lambda}_k^{-1} = O_p(T^{-2}) \]

for all \( k = 1, \ldots, m \), we have that

\[ \| D_1 \| = O_p(T^{-1}). \]

For the second term, we first show that

\[ \| \hat{D} \| = \left\| \sum_{k=m+1}^{K} T\hat{\lambda}_k^{-1} (\hat{v}_k \otimes \hat{v}_k) - \sum_{k=m+1}^{K} \lambda_k^{-1} (v_k \otimes v_k) \right\| = o_p(1). \]

Write

\[ \hat{D} = \hat{D}_1 + \hat{D}_2 \]
where

$$
\tilde{D}_1 = \sum_{k=m+1}^{K} T^{\lambda_k^{-1}} (\hat{v}_k \otimes \hat{v}_k) - \sum_{k=m+1}^{K} \lambda_k^{-1} (\hat{v}_k \otimes \hat{v}_k)
$$

and

$$
\tilde{D}_2 = \sum_{k=m+1}^{K} \lambda_k^{-1} (\hat{v}_k \otimes \hat{v}_k) - \sum_{k=m+1}^{K} \lambda_k^{-1} (v_k \otimes v_k).
$$

Since

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

by assumption we have that

$$
\frac{\log T \left( \sum_{k=m+1}^{K} \tau_k \right)^2}{T \lambda_K^2} \geq \frac{8 \log T}{T \lambda_K^4} \to 0
$$

as \( T \to \infty \). This implies that

$$
\lambda_K^{-1} = o \left( T^{1/4} \log^{-1/4} T \right). \quad (23)
$$

For large enough \( k \), we have that \( T^{-1} \hat{\lambda}_k > \lambda_k / 2 \) a.s., since if instead, then \( \left| T^{-1} \hat{\lambda}_k - \lambda_k \right| > \frac{\lambda_k}{2} \) infinitely often with positive probability, and by (23) we have that with positive probability,

$$
\limsup_{T \to \infty} T^{1/2} \log^{-1/2} T \left( \sup_{k>m} \left| T^{-1} \hat{\lambda}_k - \lambda_k \right| \right) \geq \limsup_{T \to \infty} T^{1/4} \log^{-1/4} T \left| T^{-1} \hat{\lambda}_K - \lambda_K \right| = \infty.
$$

Lemma 3.5. Specifically, for \( K \) large enough, we have that \( T^{-1} \hat{\lambda}_K > \lambda_K / 2 \) a.s. Then

$$
\left\| \tilde{D}_1 \right\| = \left\| \sum_{k=m+1}^{K} (T \hat{\lambda}_k^{-1} - \lambda_k^{-1}) (\hat{v}_k \otimes \hat{v}_k) \right\|
$$

$$
= \max_{m<k \leq K} \left| T \hat{\lambda}_k^{-1} - \lambda_k^{-1} \right|
$$

$$
\leq \sup_{k>m} \left| T^{-1} \hat{\lambda}_k - \lambda_k \right|
$$

$$
\leq \frac{2 \sup_{k>m} \left| T^{-1} \hat{\lambda}_k - \lambda_k \right|}{\lambda_K^2}.
$$
It follows from Lemma 3.5 and (23) that $\|\tilde{D}_1\| = o_p(1)$. Also,

$$
\|\tilde{D}_2\| \leq \sum_{k=m+1}^{K} \lambda_k^{-1} \|\hat{v}_k \otimes \hat{v}_k - v_k \otimes v_k\|
$$

$$
\leq \sum_{k=m+1}^{K} \lambda_k^{-1} \|\hat{v}_k \otimes (\hat{v}_k - v) + (\hat{v}_k - v) \otimes v_k\|
$$

$$
\leq 2\lambda_k^{-1} \sum_{k=m+1}^{K} \|\hat{v}_k - v_k\|
$$

$$
\leq 2\lambda_k^{-1} \|Q_{SS} - \Sigma_{SS}\| \sum_{k=m+1}^{K} \lambda_k^{-1} (\hat{v}_k \otimes \hat{v}_k)
$$

(24)

Lemma 3.5 and the assumption of this theorem, we have that $\|\tilde{D}_2\| = o_p(1)$. This completes the proof of (22).

Now

$$
D_2 = \left( \sum_{t=1}^{T} (f_t \otimes f_{t-1}) \right) \left( \sum_{k=1}^{K} \lambda_k^{-1} (\hat{v}_k \otimes \hat{v}_k) \right) \tilde{\Pi}_S - A\tilde{\Pi}_S
$$

$$
= A \left( \sum_{t=1}^{T} (f_{t-1} \otimes f_{t-1}) \right) \left( \sum_{k=m+1}^{K} \lambda_k^{-1} (\hat{v}_k \otimes \hat{v}_k) \right)
$$

$$
+ \left( \sum_{t=1}^{T} (\epsilon_t \otimes f_{t-1}) \right) \left( \sum_{k=m+1}^{K} \lambda_k^{-1} (\hat{v}_k \otimes \hat{v}_k) \right) - A\tilde{\Pi}_S
$$

$$
= \left( \sum_{t=1}^{T} (\epsilon_t \otimes f_{t-1}) \right) \left( \sum_{k=m+1}^{K} \lambda_k^{-1} (\hat{v}_k \otimes \hat{v}_k) \right) - A(\tilde{\Pi}_S - \tilde{\Pi}_{SK}),
$$

(25)

where $\tilde{\Pi}_{SK}$ is the orthogonal projection onto the space spanned by $\hat{v}_{m+1}, \ldots, \hat{v}_K$.

Write

$$
\left( \sum_{t=1}^{T} (\epsilon_t \otimes f_{t-1}) \right) \left( \sum_{k=m+1}^{K} \lambda_k^{-1} (\hat{v}_k \otimes \hat{v}_k) \right) = D_{21} + D_{22}
$$

where

$$
D_{21} = \frac{1}{T} \left( \sum_{t=1}^{T} (\epsilon_t \otimes f_{t-1}) \right) \left( \sum_{k=m+1}^{K} T\lambda_k^{-1} (\hat{v}_k \otimes \hat{v}_k) - \sum_{k=m+1}^{K} \lambda_k^{-1} (v_k \otimes v_k) \right)
$$

and

$$
D_{22} = \frac{1}{T} \left( \sum_{t=1}^{T} (\epsilon_t \otimes f_{t-1}) \right) \left( \sum_{k=m+1}^{K} \lambda_k^{-1} (v_k \otimes v_k) \right).
$$
By (19) and (22), we have that 
\[
\|D_{21}\| = o_p(1).
\]

Note that
\[
D_{22} = \frac{1}{T} \left( \sum_{t=1}^{T} (\varepsilon_t \otimes f_{t-1}^S) \right) \Pi_S \left( \sum_{k=m+1}^{K} \lambda_k^{-1} (v_k \otimes v_k) \right).
\]

Since
\[
\left\| \sum_{k=m+1}^{K} \lambda_k^{-1} (v_k \otimes v_k) \right\| = \lambda_K^{-1},
\]
by Lemma 3.6 and (23), we have that
\[
\|D_{22}\| = o_p(1).
\]

Next, we show that
\[
\left\| A\tilde{\Pi}_S - A\Pi_{SK} \right\| = o_p(1).
\] (27)

Write
\[
A\tilde{\Pi}_S - A\Pi_{SK} = A(\Pi_S - \Pi_S) + (A\Pi_S - A\Pi_{SK}) + A(\Pi_{SK} - \tilde{\Pi}_{SK})
\]
where \(\Pi_K\) is the orthogonal projection onto the subspace spanned by \(v_1, \ldots, v_K\). By Lemma 3.3 we have that
\[
\left\| A(\Pi_S - \Pi_S) \right\| = o_p(1).
\]

With a similar argument as in (24), we have that
\[
\left\| A(\Pi_{SK} - \tilde{\Pi}_{SK}) \right\| = \|A\| \left\| \sum_{k=m+1}^{K} \left[ (v_k \otimes v_k) - (\hat{v}_k \otimes \hat{v}_k) \right] \right\| = o_p(1).
\]

Let \(\tilde{A} = A\Pi_S\). Since \(A\) is compact, \(\tilde{A}^*\) is compact. Then \(\Pi_K \tilde{A}^* \to \tilde{A}^*\) in norm. To see this, notice that if instead \(\|\Pi_K \tilde{A}^* - \tilde{A}^*\| \not\to 0\), then there exists \(\epsilon > 0\) such that for any \(T\), we may find \(x_T \in H\) such that \(\|x_T\| = 1\) and that \(\|(\Pi_K(T) - 1)\tilde{A}^* x_T\| > \epsilon\). For any \(T' > T\), we have that \(\|(\Pi_K(T') - 1)\tilde{A}^* x_T'\| \geq \|(\Pi_K(T') - 1)\tilde{A}^* x_T\| > \epsilon\). Now since \(\tilde{A}^*\) is compact, there exists some subsequence \(x_{T_n}\) of \(x_T\) such that \(\tilde{A}^* x_{T_n}\) converges in norm to some \(x \in H\). Then we have that \(\|(\Pi_K(T') - 1)x\| > \epsilon\) for all \(T\). However, this is impossible since \(K(T) \to \infty\) as \(T \to \infty\). Therefore, we have that
\[
\left\| \Pi_K \tilde{A}^* - \tilde{A}^* \right\| \to 0.
\] (28)
This then implies that \( \| \hat{A} \Pi_K - \hat{A} \| \to 0 \). That is, \( \| A \Pi_K - A \Pi_S \| \to 0 \). By (25) have that
\[
\| D_2 \| = o_p(1).
\]
However, we have
\[
\| \hat{A} - A \| \leq \| D_1 \| + \| D_2 \|.
\]
from which consistency of the autoregressive estimator \( \hat{A} \) follows immediately. \( \square \)

**Proof of Theorem 3.9.** We may derive that
\[
\hat{A} - A \hat{A} = \left( \sum_{t=1}^{T} \epsilon_t \otimes f_{t-1} \right) \left( \sum_{k=1}^{K} \hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) \right).
\]
Write
\[
T(\hat{A} - A \hat{A}) \Pi_N = T \left( \sum_{t=1}^{T} \epsilon_t \otimes f_{t-1} \right) \left( \sum_{k=1}^{K} \hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) \right) \Pi_N
\]
\[= E_{11} + E_{12} + E_{13} + E_{14}, \]
where
\[
E_{11} = \left( \frac{1}{T} \sum_{t=1}^{T} \epsilon_t \otimes f_{t-1} \right) \left( \sum_{k=1}^{m} T^2 \hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) \right) (\Pi_N - \hat{\Pi}_N),
\]
\[
E_{12} = \left( \sum_{t=1}^{T} \epsilon_t \otimes f_{t-1} \right) (\hat{\Pi}_S - \Pi_S) \left( \sum_{k=m+1}^{K} T \hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) \right) (\Pi_N - \hat{\Pi}_N),
\]
\[
E_{13} = \left( \sum_{t=1}^{T} \epsilon_t \otimes f_{t-1} \right) \Pi_S \left( \sum_{k=m+1}^{K} T \hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) \right) (\Pi_N - \hat{\Pi}_N),
\]
and
\[
E_{14} = \left( \frac{1}{T} \sum_{t=1}^{T} \epsilon_t \otimes f_{t-1} \right) \left( \sum_{k=1}^{K} T^2 \hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) \right) \hat{\Pi}_N.
\]
By (19), Lemma 3.3 and Theorem 3.4, we have that \( \| E_{11} \| = O_p(T^{-1}) \). By (19), (22), (23), Lemma 3.3 and Theorem 3.4, we have that \( \| E_{12} \| = o_p(T^{-3/4} \log^{-1/4} T) \). Note that
\[
\left( \sum_{t=1}^{T} \epsilon_t \otimes f_{t-1} \right) \Pi_S = \sum_{t=1}^{T} \epsilon_t \otimes f_{t-1}^S,
\]
then by, (22), (23), Lemma 3.3, Lemma 3.6 and Theorem 3.4, we have that \( \| E_{13} \| = \)
o_p(T^{-1/4} \log^{-1/4} T). Now again by (19), (22) and Theorem 3.4 we have that

$$E_{14} = \left( \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t \otimes f_{t-1} \right) \left( \sum_{k=1}^{K} T^2 \hat{\lambda}_k^{-1}(\hat{u}_k \otimes \hat{v}_k) \right) \Pi_N$$

$$\to_d \left( \int_0^1 (dW \otimes W_P)(r) \right) \left( \sum_{k=1}^{m} \lambda_k^{-1}(v_k \otimes v_k) \right)$$

$$= \left( \int_0^1 (dW \otimes W_P)(r) \right) \left( \int_0^1 (W_P \otimes W_P)(r) dr \right)^{-1}.$$

Then we have

$$T(\hat{A} - A\hat{\Pi}_K)\Pi_N \to_d \left( \int_0^1 (dW \otimes W_P)(r) \right) \left( \int_0^1 (W_P \otimes W_P)(r) dr \right)^{-1}.$$

\[ \square \]

**Proof of Theorem 3.10.** For any $v \notin H_N$, we have that

$$(\hat{A} - A\hat{\Pi}_K)v = E_{21} + E_{22} + E_{23} + E_{24}$$

where

$$E_{21} = \left( \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t \otimes f_{t-1} \right) \left( \sum_{k=1}^{K} T^2 \hat{\lambda}_k^{-1}(\hat{u}_k \otimes \hat{v}_k) \right) (\Pi_S - \hat{\Pi}_S)v,$$

$$E_{22} = \left( \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t \otimes f_{t-1} \right) (\hat{\Pi}_S - \Pi_S) \left( \sum_{k=1}^{K} T^2 \hat{\lambda}_k^{-1}(\hat{u}_k \otimes \hat{v}_k) \right) \hat{\Pi}_Sv,$$

$$E_{23} = \frac{1}{\sqrt{T}} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t \otimes f_{t-1}^S \right) \left( \sum_{k=m+1}^{K} \hat{T} \lambda_k^{-1}(\hat{u}_k \otimes \hat{v}_k) - \sum_{k=m+1}^{K} \lambda_k^{-1}(v_k \otimes v_k) \right) v,$$

and

$$E_{24} = \frac{1}{\sqrt{T}} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t \otimes f_{t-1}^S \right) \left( \sum_{k=m+1}^{K} \lambda_k^{-1}(v_k \otimes v_k) \right) v.$$

By (19), (20), (22), (23), Lemma 3.3 and Theorem 3.4, we have that $\|E_{21}\| = o_p(T^{-3/4} \log^{-1/4} T)$. Similarly, $\|E_{22}\| = o_p(T^{-3/4} \log^{-1/4} T)$. By Lemma 3.6 and (22), we have that $\|E_{23}\| = o_p(T^{-1/2})$. 

48
Note that

\[ Z_t = (\varepsilon_t \otimes f^S_{t-1}) \left( \sum_{k=m+1}^{K} \lambda_k^{-1}(v_k \otimes v_k) \right) v \]

\[ = \left\langle f^S_{t-1}, \left( \sum_{k=m+1}^{K} \lambda_k^{-1}(v_k \otimes v_k) \right) v \right\rangle \varepsilon_t \]

is a martingale difference sequence with respect to \( \mathcal{F}_t = \sigma(\varepsilon_i : i \leq t) \). Since \( f^S_{t-1} \) is independent of \( \varepsilon_t \), we have that

\[ E(Z_t \otimes Z_t) = E \left( f^S_{t-1}, \left( \sum_{k=m+1}^{K} \lambda_k^{-1}(v_k \otimes v_k) \right) v \right)^2 E(\varepsilon_t \otimes \varepsilon_t) \]

\[ = \left\langle \left( \sum_{k=m+1}^{K} \lambda_k^{-1}(v_k \otimes v_k) \right) v, E(f^S_{t-1} \otimes f^S_{t-1}) \left( \sum_{k=m+1}^{K} \lambda_k^{-1}(v_k \otimes v_k) \right) v \right\rangle \Sigma \]

\[ = \left\langle \left( \sum_{k=m+1}^{K} \lambda_k^{-1}(v_k \otimes v_k) \right) v, \left( \sum_{k=m+1}^{K} (v_k \otimes v_k) \right) v \right\rangle \Sigma \]

\[ = \left( \sum_{k=m+1}^{K} \lambda_k^{-1}(v_k, v)^2 \right) \Sigma. \]

By the central limit theorem for real-valued martingale difference sequence, we have that

for any \( x \in H \),

\[ \frac{1}{\sqrt{TS_K(v)}} \sum_{t=1}^{T} \langle x, Z_t \rangle \to_d N(0, \langle x, \Sigma x \rangle) \tag{29} \]

where \( S_K(x) = \sum_{k=m+1}^{K} \lambda_k^{-1}(v_k, v)^2 \). Next, we show that the sequence

\[ \tilde{Z}_T = \frac{1}{\sqrt{TS_K(v)}} \sum_{t=1}^{T} Z_t \]

is tight. Let \( \Pi_n^\Sigma \) be the orthogonal projection onto the space spanned by the first \( n \) eigenvectors of the variance operator \( \Sigma \). Since that for any \( \epsilon > 0 \),

\[ \mathbb{P} \left( \left\| (1 - \Pi_n^\Sigma) \tilde{Z}_T \right\| > \epsilon \right) \leq \frac{E \left\| (1 - \Pi_n^\Sigma) \tilde{Z}_T \right\|^2}{\epsilon^2} \]
and that

\[
\mathbb{E} \left\| (1 - \Pi_n) \tilde{Z}_T \right\|^2 = \frac{1}{TS_K(v)} \mathbb{E} \left( \left\| \sum_{t=1}^T \left( \sum_{k=m+1}^K \lambda_k^{-1}(v_k \otimes v_k) f_{i_{t-1},v}^S \right) (1 - \Pi_n^\Sigma) \varepsilon_t \right\|^2 \right)
\]

\[
= \frac{1}{TS_K(v)} \sum_{t=1}^T \mathbb{E} \left( \left\| \left( \sum_{k=m+1}^K \lambda_k^{-1}(v_k \otimes v_k) f_{i_{t-1},v}^S \right) \right\|^2 (1 - \Pi_n^\Sigma) \varepsilon_t^2 \right)
\]

\[
= \frac{1}{SK(v)} \mathbb{E} \left\| (1 - \Pi_n^\Sigma) \varepsilon_t \right\|^2 \mathbb{E} \left( \left( \sum_{k=m+1}^K \lambda_k^{-1}(v_k \otimes v_k) f_{i_{t-1},v}^S \right) \right)^2
\]

\[
= \text{tr} \left( (1 - \Pi_n^\Sigma) \Sigma \right)
\]

\[
\rightarrow 0
\]

as \( n \to \infty \), we have

\[
\lim_{n \to \infty} \sup_T \mathbb{P} \left( \left\| (1 - \Pi_n^\Sigma) \tilde{Z}_T \right\| > \epsilon \right) = 0.
\]

This implies that \( (\tilde{Z}_t) \) is tight, so the central limit theorem for the real valued martingale difference sequence as in (29) implies a central limit theorem for the \( H \)-valued martingale difference sequence \( Z_t \):

\[
\sqrt{\frac{T}{SK(x)}} E_{24} = \frac{1}{\sqrt{TS_K(v)}} \sum_{t=1}^T Z_t \to_d \mathbb{N}(0, \Sigma).
\]

This completes the proof of

\[
\sqrt{\frac{T}{SK(x)}}(\hat{\Lambda} - A\Pi_K) v \to_d \mathbb{N}(0, \Sigma)
\]

for any \( x \notin H_N \). \( \Box \)

**Proof of Theorem 3.11.** Since

\[
\left\| A(\Pi_K - A)\Pi_N \right\| \leq \left\| A(\Pi_K - \Pi_K)\Pi_N \right\| + \left\| A(\Pi_K - 1)\Pi_N \right\|
\]

\[
= \left\| A(\Pi_K - \Pi_K)\Pi_N \right\|
\]

\[
\leq \left\| A \right\| \left\| \Pi_K - \Pi_K \right\|,
\]

50
to show that \((A\hat{\Pi}_K - A)\Pi_N = o_p(T^{-1/2}K^{1/2})\), it suffices to show that
\[
\left\| \hat{\Pi}_K - \Pi_K \right\| = o_p(T^{-1/2}K^{1/2}).
\]

Write
\[
\left\| \hat{\Pi}_K - \Pi_K \right\| = \left\| \hat{\Pi}_{NK} - \Pi_{NK} - \Pi_{SK} \right\| \leq \left\| \hat{\Pi}_{NK} - \Pi_{NK} \right\| + \left\| \hat{\Pi}_{SK} - \Pi_{SK} \right\|
\]
where \(\hat{\Pi}_{NK}\) and \(\Pi_{NK}\) are the orthogonal projections onto the space spanned by \(\hat{v}_1, \ldots, \hat{v}_m\) and onto the space spanned by \(v_1, \ldots, v_m\) respectively, and \(\hat{\Pi}_{SK}\) and \(\Pi_{SK}\) are the orthogonal projections onto the space spanned by \(\hat{v}_{m+1}, \ldots, \hat{v}_K\) and onto the space spanned by \(v_{m+1}, \ldots, v_K\) respectively. Since
\[
\left\| \hat{\Pi}_{NK} - \Pi_{NK} \right\| = \left\| \hat{\Pi}_N - \Pi_N \right\| = O_p(T^{-1}),
\]
it suffices to show that
\[
\left\| \hat{\Pi}_{SK} - \Pi_{SK} \right\| = o_p(T^{-1/2}K^{1/2}).
\]

Let \(K_m = K - m\). Subsequently, we write \(\hat{v}_{m+k} = \hat{v}_k^S\) and \(v_{m+k} = v_k^S\) for \(k = 1, \ldots, K_m\). Moreover, we let \((\tilde{v}_k^S)\) be the eigenvectors associated with the nonzero eigenvalues \((\tilde{\lambda}_k^S)\) of \(Q_{SS}\) such that \(\tilde{\lambda}_1^S \geq \tilde{\lambda}_2^S \geq \cdots\), and define \(\hat{\Pi}_{SK}\) to be the orthogonal projection on the subspace of \(H\) spanned by \((\tilde{v}_k^S)\) for \(k = 1, \ldots, K_m\). Note that \(\tilde{v}_k^S = v_k(Q_{SS})\) and \(v_k^S = v_k(\Sigma_{SS})\) for \(k = 1, 2, \ldots\).

Write
\[
\left\| \hat{\Pi}_{SK} - \Pi_{SK} \right\| \leq \left\| \hat{\Pi}_{SK} - \hat{\Pi}_{SK} \right\| + \left\| \hat{\Pi}_{SK} - \Pi_{SK} \right\|. \tag{30}
\]
For the first term in (30), it follows from Hu, Park and Qian (2016) that under Assumption 3.2,
\[
\left\| \hat{\Pi}_{SK} - \Pi_{SK} \right\| = o_p(T^{-1/2}K^{1/2}). \tag{31}
\]
To analyze the second term in (30), we note that \((\tilde{v}_k^S)\) and \((\tilde{v}_k^S)\) are the eigenvectors associated with the leading eigenvalues of
\[
\hat{\Pi}_S Q_{S} \hat{\Pi}_S = \hat{\Pi}_S \left( \frac{1}{T} \sum_{t=1}^{T} (f_{t-1} \otimes f_{t-1}) \right) \hat{\Pi}_S
\]
and
\[
Q_{SS} = \Pi_S \left( \frac{1}{T} \sum_{t=1}^{T} (f_{t-1} \otimes f_{t-1}) \right) \Pi_S,
\]

51
respectively.

It follows from Lemma 3.3 that

\[ \hat{\Pi}_S \Pi_N = O_p(T^{-1}) \quad \text{and} \quad \Pi_N \hat{\Pi}_S = O_p(T^{-1}), \]

and we have

\[ \Pi_N \left( \frac{1}{T^2} \sum_{t=1}^{T} (f_{t-1} \otimes f_{t-1}) \right) \Pi_N = O_p(1), \]

\[ \Pi_S \left( \frac{1}{T} \sum_{t=1}^{T} (f_{t-1} \otimes f_{t-1}) \right) \Pi_N = O_p(1), \]

and

\[ \Pi_N \left( \frac{1}{T} \sum_{t=1}^{T} (f_{t-1} \otimes f_{t-1}) \right) \Pi_S = O_p(1). \]

Therefore,

\[ \hat{\Pi}_S Q \hat{\Pi}_S = \hat{\Pi}_S Q S \hat{\Pi}_S + O_p(T^{-1}) = Q S S + O_p(T^{-1}) \]

uniformly in \( T \), and consequently,

\[ \max_{1 \leq k \leq K_m} \| \hat{v}_k^S - \tilde{v}_k^S \| = O_p(T^{-1}), \]

from which it follows that

\[ \| \hat{\Pi}_{SK} - \tilde{\Pi}_{SK} \| = O_p(T^{-1} K). \quad (32) \]

Now it follows from (30), (31) and (32) that

\[ \| \hat{\Pi}_{SK} - \Pi_{SK} \| = o_p(T^{-1/2} K^{1/2}), \]

which shows that \( (A \hat{\Pi}_K - A) \Pi_N = o_p(T^{-1/2} K^{1/2}) \).

Also,

\[ \| A(\hat{\Pi}_K - \Pi_K) S v \| \leq \| A(\hat{\Pi}_K - \Pi_K) S v \| + \| A(\Pi_K - 1) S v \| \]

\[ \leq \| A \| \| \hat{\Pi}_K - \Pi_K \| \| v \| + \| A \| \| \Pi_K - 1 \| \| v \| \]

\[ = o_p(T^{-1/2} K^{1/2}) + O(\| (1 - \Pi_K) v \|), \]

and the proof is complete.
Proof of Lemma 3.12. Write

\[ \hat{A}f_T - A\hat{\Pi}_K f_T = D_3 + D_4 + D_5 \]

where

\[ D_3 = (\hat{A} - A\hat{\Pi}_K)\hat{\Pi}_N f_T, \]
\[ D_4 = (\hat{A} - A\hat{\Pi}_K)(\hat{\Pi}_S - \Pi_S)f_T, \]

and

\[ D_5 = (\hat{A} - A\hat{\Pi}_K)\Pi_S f_T. \]

Note that

\[ D_3 = D_1 f_T, \]

then by Lemma 3.1 and (21) we have that

\[ \|D_3\| = O_p(T^{-1/2}). \]

Since

\[ \|D_4\| \leq \|\hat{A} - A\| \|\hat{\Pi}_K\| \|\hat{\Pi}_S - \Pi_S\| \|f_T\|, \]

by Lemma 3.1, Lemma 3.3 and the consistency of \( \hat{A} \), we have that \( \|D_4\| = o_p(T^{-1/2}) \).

Following the proof of Theorem 3.10, we have that \( D_5 = \tilde{D}_5 + o_p(T^{-1/2}) \) where

\[ \tilde{D}_5 = \frac{1}{T} \left( \sum_{t=1}^{T} (\varepsilon_t \otimes f_{t-1}^S) \right) \left( \sum_{k=m+1}^{K} \lambda_k^{-1} (v_k \otimes v_k) \right) f_T^S. \]

Therefore, it suffices to show that

\[ \sqrt{T/K} \tilde{D}_5 \rightarrow_d N(0, \Sigma). \] (33)

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

\[ \sqrt{\frac{1}{TK}} \left( \sum_{t=1}^{T} (\varepsilon_t \otimes f_{t-1}^S) \right) \left( \sum_{k=m+1}^{K} \lambda_k^{-1} (v_k \otimes v_k) \right) f_T^S \rightarrow_d N(0, \Sigma). \] (34)

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

53
operator, we have that
\[ f_t^S = \Pi_S(Af_{t-1} + \varepsilon_t) \]
\[ = \Pi_S(A(f_{t-1}^N + f_{t-1}^S)) + \varepsilon_t^S \]
\[ = \Pi_S\Pi_Sf_{t-1}^S + \varepsilon_t^S. \]

This implies that \((f_t^S)\) has a functional autoregressive representation with autoregressive operator \(\Pi_S\Pi_S\). For convenience, let \(\tilde{A} = \Pi_S\Pi_S\). Since \(\|\tilde{A}\| \leq \|A\|\), the first order difference equation \(g_t = \tilde{A}g_{t-1} + \varepsilon_t^S\) 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 \(D_{54}\) as
\[
\left( \sum_{t=1}^{T} (\varepsilon_t \otimes f_{t-1}^S) \right) \left( \sum_{k=m+1}^{K} \lambda_t^{-1}(v_i \otimes v_i) \right) f_{T+1}^S
\]
\[ = \sum_{t=1}^{T} \langle f_{t-1}^S, Q_2 f_{T+1}^S \rangle \varepsilon_t
\]
\[ = \sum_{t=1}^{T} \langle Q_2 f_{t-1}^S, f_{T+1}^S \rangle \varepsilon_t
\]
\[ = \sum_{t=1}^{T} (Z_t^+ + Z_t^0 + Z_t^-)
\]
where
\[ Z_t^+ = \langle Q_2 f_{t-1}^S, f_{t+1}^S \rangle \varepsilon_t, \]
\[ Z_t^0 = \langle Q_2 f_{t-1}^S, (\tilde{A})_{T+1-t} \varepsilon_t^S \rangle \varepsilon_t, \]
\[ Z_t^- = \langle Q_2 f_{t-1}^S, (\tilde{A})_{T+2-t} f_{t-1}^S \rangle \varepsilon_t, \]
and
\[ f_{t+} = \varepsilon_{T+1}^S + \tilde{A}\varepsilon_t^S + \cdots + (\tilde{A})_{T-t}^S \varepsilon_{t+1}^S. \]

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 \(F_t\). Following the proof of Lemma 5.8 in Mas (2007), it is easy to show that
\[
\mathbb{E}(Z_t^+ \otimes Z_n^+) = 0
\]
for \( t < s \), and

\[
\mathbb{E}(Z_t^+ \otimes Z_t^+) = \mathbb{E}(Q_2^+ f_{l-1}^S, f_{l+1}^S)^2 \Sigma.
\]

Note that \( Q_2^+ \) is the inverse of \( \mathbb{E}(f_{l-1}^S \otimes f_{l+1}^S) \) restricted on the span of \( v_{m+1}, \ldots, v_K \), we may follow the proof of Lemma 5.8 in Mas (2007) to obtain that

\[
\mathbb{E}(Z_t^+ \otimes Z_t^+) = (K - m - \text{tr} \left( Q_2^+ (\hat{A})^{T-t+1} + \Sigma_{SS}(\hat{A}^*)^{T-t+1} \right)) \Sigma.
\]

Since \( \left| \text{tr} \left( Q_2^+ (\hat{A})^{T-t+1} + \Sigma_{SS}(\hat{A}^*)^{T-t+1} \right) \right| \) is bounded by a constant under the assumption in the theorem, we have that \( \mathbb{E}(Q_2^+ f_{l-1}^S, f_{l+1}^S)^2 \sim TK \). 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

\[
\frac{1}{\sqrt{TK}} \sum_{t=1}^T (Z_t^+, x) \rightarrow_d \mathbb{N}(0, \langle x, \Sigma x \rangle).
\] (35)

We may follow the proof of Lemma 5.9 in Mas (2007) to show that \( \frac{1}{\sqrt{TK}} \sum_{t=1}^T Z_t^+ \) is a tight sequence. This then implies that it converges, and the limiting distribution is fully characterized by the finite distribution on the right hand side of (35). That is, we have that

\[
\frac{1}{\sqrt{TK}} \sum_{t=1}^T Z_t^+ \rightarrow_d \mathbb{N}(0, \Sigma).
\]

One may follow Lemma 5.10 in Mas (2007) to show that

\[
\frac{1}{\sqrt{TK}} \sum_{t=1}^T Z_0^+ \rightarrow_p 0
\]

and that

\[
\frac{1}{\sqrt{TK}} \sum_{t=1}^T Z_t^- \rightarrow_p 0.
\]

The conclusion then follows immediately.

**Proof of Theorem 3.13.** In view of Lemma 3.12, it suffices to show that

\[
\sqrt{T/K}(A\hat{\Pi}_K f_T - Af_T) = o_p(1).
\]

Notice that

\[
A\hat{\Pi}_K f_T - Af_T = (A\hat{\Pi}_K \hat{\Pi}_N f_T - A\hat{\Pi}_N f_T) + (A\hat{\Pi}_K \hat{\Pi}_S f_T - A\hat{\Pi}_S f_T)
\]

55
and that
\[ A\hat{\Pi}_K \hat{\Pi}_N f_T - A\hat{\Pi}_N f_T = A\hat{\Pi}_N f_T - A\hat{\Pi}_N f_T = 0, \]
it then suffices to show that
\[ \sqrt{T/K}(A\hat{\Pi}_K \hat{\Pi}_S f_T - A\hat{\Pi}_S f_T) = o_p(1). \]

Since
\[ A\hat{\Pi}_K \hat{\Pi}_S f_T - A\hat{\Pi}_S f_T = (A\hat{\Pi}_K \Pi_S - A\Pi_S) f_T + (A\hat{\Pi}_K - A)(\hat{\Pi}_S - \Pi_S) f_T, \]
and that \[ \|A\hat{\Pi}_K - A\| \leq \|A\|, \]
by Lemma 3.1 and Lemma 3.3, it suffices to show that
\[ \left\| (A\hat{\Pi}_K \Pi_S - A\Pi_S) f_T \right\| = \left\| (A\hat{\Pi}_K - A) f_T^S \right\| = o_p(T^{-1/2}K^{1/2}). \]

Write
\[ (A\hat{\Pi}_K - A) f_T^S = A(\hat{\Pi}_K - \Pi_K) f_T^S + A(\Pi_K - 1) f_T^S. \]
In the proof of Theorem 3.11, we have shown that
\[ \left\| \hat{\Pi}_K - \Pi_K \right\| = o_p(T^{-1/2}K^{1/2}) \]
under Assumption 3.2. Also,
\[ \mathbb{E} \left\| (\Pi_K - 1) f_T^S \right\|^2 = \mathbb{E} \left\| \sum_{k=K+1}^{\infty} \langle v_k, f_T^S \rangle v_k \right\|^2 \]
\[ = \mathbb{E} \left\| \sum_{k=K+1}^{\infty} \langle v_k, \hat{f}_T \rangle v_k \right\|^2 \]
\[ = \mathbb{E} \left( \sum_{k=K+1}^{\infty} \langle v_k, \hat{f}_T \rangle^2 \right) \]
\[ = \sum_{k=K+1}^{\infty} \lambda_k \]
\[ = o(T^{-1} K) \]
by Assumption 3.4, which implies that
\[ \left\| (\Pi_K - 1) f_T^S \right\| = o_p(T^{-1/2}K^{1/2}). \]
This then completes the proof. 

**Proof of Theorem 4.2.** Note that we have

\[ f_t^\mu = A f_{t-1}^\mu + \varepsilon_t + (A - 1) \left( \frac{1}{T} \sum_{t=1}^{T} f_t - \mu \right) \]

\[ = A f_{t-1}^\mu + \varepsilon_t + (A - 1) \left( \frac{1}{T} \sum_{t=1}^{T} f_t^T - \mu \right), \]

where the second equality follows from the fact that \( A \) restricted on \( H_N \) is the identity operator. Now write

\[ \hat{A} = D_{71} + D_{72} + D_{73} + D_{74}, \]

where

\[ D_{71} = \left( \sum_{t=1}^{T} (Af_{t-1}^\mu + \varepsilon_t) \otimes f_{t-1}^\mu \right) \left( \sum_{k=1}^{K} \hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) \right), \]

\[ D_{72} = (A - 1) \left[ \left( \frac{1}{T} \sum_{t=1}^{T} f_t^T - \mu \right) \otimes \left( \frac{1}{T} \sum_{t=1}^{T} f_{t-1}^{\mu_S} \right) \right] \left( \sum_{k=1}^{K} T \hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) \right), \]

\[ D_{73} = (A - 1) \left[ \left( \frac{1}{T} \sum_{t=1}^{T} f_t^T - \mu \right) \otimes \left( \frac{1}{T^2} \sum_{t=1}^{T} f_{t-1}^{\mu_N} \right) \right] \left( \sum_{k=1}^{m} T^2 \hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) \right), \]

and

\[ D_{74} = (A - 1) \left[ \left( \frac{1}{T} \sum_{t=1}^{T} f_t^T - \mu \right) \otimes \left( \frac{1}{T} \sum_{t=1}^{T} f_{t-1}^{\mu_N} \right) \right] \cdot \left( \sum_{k=m+1}^{K} T \hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) - \sum_{k=m+1}^{K} \hat{\lambda}_k^{-1}(v_k \otimes v_k) \right), \]

where in turn

\[ f_t^{\mu_S} = f_t^S - \frac{1}{T} \sum_{t=1}^{T} f_t^S \quad \text{and} \quad f_t^{\mu_N} = f_t^N - \frac{1}{T} \sum_{t=1}^{T} f_t^N. \]

By Theorem 3.10 in Bosq (2000), we have that

\[ \frac{1}{T} \sum_{t=1}^{T} f_t^T - \mu = O_p(1/\sqrt{T}). \]
Also,
\[
\sum_{t=1}^{T} f_{t-1}^u = \sum_{t=1}^{T} \left( f_{t-1} - \frac{1}{T} \sum_{t=1}^{T} f_t \right) = f_0 - f_T,
\]
then we have that \( \left\| \sum_{t=1}^{T} f_{t-1}^\mu \right\| = O_p(1) \), and \( \left\| \sum_{t=1}^{T} f_{t-1}^N \right\| = O_p(\sqrt{T}) \). Also, we have that
\[
\left\| \sum_{k=1}^{K} T \hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) \right\| = o_p(T^{1/4} \log^{-1/4} T),
\]
\[
\left\| \sum_{k=1}^{m} T^2 \hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) \right\| = O_p(1)
\]
and
\[
\left\| \sum_{k=m+1}^{K} T \hat{\lambda}_k^{-1}(\hat{v}_k \otimes \hat{v}_k) - \sum_{k=m+1}^{K} \lambda_k^{-1}(v_k \otimes v_k) \right\| = o_p(1).
\]
Then we have \( \| D_{72} \| = o_p(T^{-5/4} \log^{-1/4} T), \| D_{73} \| = O_p(T^{-2}), \) and \( \| D_{74} \| = O_p(T^{-1}) \). This implies that we may write
\[
\hat{A} = D_{71} + O_p(T^{-1}). \tag{36}
\]

A careful examination of the proof of Theorem 3.7 suggests that without any essential change in the proof we may show that \( \| D_{71} - A \| \to_p 0 \). This then establishes the consistency. To establish the asymptotic normality, we first show that
\[
\sqrt{T/K/}(\hat{A} f_T^\mu - A f_T^\mu) \to_d N(0, \Sigma) \tag{37}
\]
and
\[
\sqrt{T/K}(\hat{A} f_T^\mu - A f_T^\mu) \to_d N(0, \Sigma) \tag{38}
\]
under the assumptions for Lemma 3.12 and Theorem 3.13 respectively. A careful examination of the proof for Lemma 3.12 indicate that with the presence of (36), to prove (37), it suffices to prove a counterpart of (34), namely
\[
\sqrt{1/TK} \left( \sum_{t=1}^{T} (\varepsilon_t \otimes f_{t-1}^S) \right) \left( \sum_{k=m+1}^{K} \lambda_k^{-1}(v_k \otimes v_k) \right) f_{T+1}^S \to_d N(0, \Sigma).
\]
We may write the left hand side of the above equation as

\[
\sqrt{\frac{1}{TK}} \left( \sum_{t=1}^{T} (\varepsilon_t \otimes ((f_{t-1}^S - \mu^S) + (\mu^S - \frac{1}{T} \sum_{t=1}^{T} f_t^S))) \right) \\
\cdot \left( \sum_{k=m+1}^{K} \lambda_k^{-1}(v_k \otimes v_k) \right) \left( (f_{t-1}^S - \mu^S) - \mu^S - \frac{1}{T} \sum_{t=1}^{T} f_t^S \right)
\]

where \(\mu^S = \Pi_{S\mu} \). Notice that we have \(\left\| \mu^S - T^{-1} \sum_{t=1}^{T} f_t^S \right\| = O_p(1/\sqrt{T})\) and that

\[
\left\| \sum_{t=1}^{T} \varepsilon_t \otimes \left( \mu^S - \frac{1}{T} \sum_{t=1}^{T} f_t^S \right) \right\| \leq \left\| \mu^S - \frac{1}{T} \sum_{t=1}^{T} f_t^S \right\| \left\| \sum_{t=1}^{T} \varepsilon_t \right\| = O_p(1),
\]

it is easy to show that

\[
\sqrt{\frac{1}{TK}} \left( \sum_{t=1}^{T} (\varepsilon_t \otimes f_{t-1}^S) \right) \left( \sum_{k=m+1}^{K} \lambda_k^{-1}(v_k \otimes v_k) \right) f_{T+1}^S
\]

\[
= \sqrt{\frac{1}{TK}} \left( \sum_{t=1}^{T} (\varepsilon_t \otimes (f_{t-1}^S - \mu^S)) \right) \left( \sum_{k=m+1}^{K} \lambda_k^{-1}(v_k \otimes v_k) \right) (f_{t-1}^S - \mu^S) + o_p(1).
\]

Without any essential changes in the proof for (34), we may prove that

\[
\sqrt{\frac{1}{TK}} \left( \sum_{t=1}^{T} (\varepsilon_t \otimes (f_{t-1}^S - \mu^S)) \right) \left( \sum_{k=m+1}^{K} \lambda_k^{-1}(v_k \otimes v_k) \right) (f_{t-1}^S - \mu^S) \to_d N(0, \Sigma).
\]

This then establishes (37).

For the proof of (38) we may follow the proof of Theorem 3.13 essentially with \(f\) replaced by \(f^\mu\) at corresponding places. The only thing we need to take special case is that in this case we need to redefine \(Q_{SS}\) and \(\hat{Q}_{SS}\) as

\[
Q_{SS} = \frac{1}{T} \sum_{t=1}^{T} (f_{t-1}^S - \mu^S) \otimes (f_{t-1}^S - \mu^S)
\]

and

\[
\hat{Q}_{SS} = \frac{1}{T} \sum_{t=1}^{T} \hat{\Pi}_S f_{t-1}^\mu \otimes \hat{\Pi}_S f_{t-1}^\mu.
\]

Once we reestablish that \(\left\| \hat{Q}_{SS} - Q_{SS} \right\| = O_p(T^{-1})\), everything follows essentially and we
may establish (38). In fact, write

\[ \hat{Q}_{SS} - Q_{SS} = D_{81} + D_{82} \]

where

\[ D_{81} = \frac{1}{T} \sum_{t=1}^{T} \hat{\Pi}_S f^\mu_{t-1} \otimes \hat{\Pi}_S f^\mu_{t-1} - \frac{1}{T} \sum_{t=1}^{T} \Pi_S f^\mu_{t-1} \otimes \Pi_S f^\mu_{t-1} \]

and

\[ D_{82} = \frac{1}{T} \sum_{t=1}^{T} \Pi_S f^\mu_{t-1} \otimes \Pi_S f^\mu_{t-1} - \frac{1}{T} \sum_{t=1}^{T} \Pi_S (f_{t-1} - \mu) \otimes \Pi_S (f_{t-1} - \mu). \]

Following the proof of Theorem 3.13, we have that \( \|D_{81}\| = O_p(T^{-1}) \). Also notice that

\[ \Pi_S f^\mu_{t-1} - \Pi_S (f_{t-1} - \mu) = \mu^S - \sum_{t=1}^{T} f^S_t = O_p(1/\sqrt{T}), \]

we may show that \( \|D_{82}\| = O_p(T^{-1}) \). This then established the desired equality and therefore established (38). Having established (38), it is easy to establish the asymptotic normality by noticing that

\[ (\hat{f}_{T+1} - \mu) - A(f_T - \mu) = \hat{A}f^\mu_T - Af^\mu_T + (1 - A) \left( \frac{1}{T} \sum_{t=1}^{T} f^T_t - \mu \right), \]

where we have used the fact that \( A \) restricted on \( H_N \) is the identity operator, and that

\[ \frac{1}{T} \sum_{t=1}^{T} f^T_t - \mu = O_p(1/\sqrt{T}). \]

The proof is therefore complete.

Proof of Theorem 3.14. Let \( \alpha(A) \) be the set of eigenvalues of \( A \) except for the unit eigenvalues. Since \( \hat{A} \) is consistent, by Theorem 4.1 and 4.2 in Gohberg et al. (1990), for \( T \) large enough, there exists a contour \( \Gamma \) around \( \alpha(A) \) such that \( \Gamma \) separates \( \alpha(A) \) from \( \{1, \hat{\alpha}_1, \ldots, \hat{\alpha}_m\} \), and also separates \( \{\hat{\alpha}_{m+1}, \ldots, \hat{\alpha}_K\} \) from \( \{1, \hat{\alpha}_1, \ldots, \hat{\alpha}_m\} \).

Note that by Theorem 2.1, \( H_T \) is the invariant subspace of \( A \) corresponding to the unit eigenvalue, and that \( H_T \) is the invariant subspace of \( A \) corresponding to \( \alpha(A) \). By Theorem 2.2 in Gohberg et al. (1990), we have that

\[ \Pi_T = \frac{1}{2\pi i} \oint_{\Gamma} (\lambda - A)^{-1} d\lambda. \]

\[ (39) \]
Also, by Theorem 7.17 and Theorem 7.24 in Friedberg et al. (1989), \( \hat{H}_P \) and \( \hat{H}_T \) are two invariant subspaces of \( \hat{A} \) corresponding to the eigenvalues \( \hat{\alpha}_1, \ldots, \hat{\alpha}_m \) and \( \hat{\alpha}_{m+1}, \ldots, \hat{\alpha}_K \) respectively. Then

\[
\hat{\Pi}_T = \frac{1}{2\pi i} \oint_{\Gamma} (\lambda - \hat{A}_T)^{-1} d\lambda.
\]

By (39) and the continuity of analytic functions, we have that

\[
\|\hat{\Pi}_T - \Pi_T\| = o_p(1).
\]

Similarly one may show that

\[
\|\hat{\Pi}_P - \Pi_P\| = o_p(1).
\]
Appendix B  Factor Analysis Data Description

This appendix lists all the FRED-MD database variables except for the variable ACOGNO (new orders for consumer goods). We construct three sets of candidate variables for the adaptive LASSO analysis. Candidate set A contains all 127 variables listed in the following table. Candidate set B contains all but the interest rate variables. Candidate set C contains selected variables that are used in the analysis in the main text. Column 3 of the table below gives information on whether a variable is included in the candidate set A, B, and/or C. A “y” indicates that the variable is included in the corresponding candidate set.

Table 3: Data Description

<table>
<thead>
<tr>
<th>No.</th>
<th>Mnemonic</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RPI</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>real personal income</td>
</tr>
<tr>
<td>2</td>
<td>W875RX1</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>real personal income exclude current transfer receipts</td>
</tr>
<tr>
<td>3</td>
<td>DPCERA3M086SBEA</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>real personal consumption expenditures</td>
</tr>
<tr>
<td>4</td>
<td>CMRMTSPLx</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>real manufacturing and trade industries sales</td>
</tr>
<tr>
<td>5</td>
<td>RETAILx</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>retail and food services sales</td>
</tr>
<tr>
<td>6</td>
<td>INDPRO</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>industrial production index (IP)</td>
</tr>
<tr>
<td>7</td>
<td>IPFPNSS</td>
<td>y</td>
<td>y</td>
<td></td>
<td>IP: final products and nonindustrial supplies</td>
</tr>
<tr>
<td>8</td>
<td>IPFINAL</td>
<td>y</td>
<td>y</td>
<td></td>
<td>IP: final products (market group)</td>
</tr>
<tr>
<td>9</td>
<td>IPCONGD</td>
<td>y</td>
<td>y</td>
<td></td>
<td>IP: consumer goods</td>
</tr>
<tr>
<td>10</td>
<td>IPDCONGD</td>
<td>y</td>
<td>y</td>
<td></td>
<td>IP: durable consumer goods</td>
</tr>
<tr>
<td>11</td>
<td>IPNCONGD</td>
<td>y</td>
<td>y</td>
<td></td>
<td>IP: nondurable consumer goods</td>
</tr>
<tr>
<td>12</td>
<td>IPBUSEQ</td>
<td>y</td>
<td>y</td>
<td></td>
<td>IP: business equipment</td>
</tr>
<tr>
<td>13</td>
<td>IPMAT</td>
<td>y</td>
<td>y</td>
<td></td>
<td>IP: materials</td>
</tr>
<tr>
<td>14</td>
<td>IPDMAT</td>
<td>y</td>
<td>y</td>
<td></td>
<td>IP: durable materials</td>
</tr>
<tr>
<td>15</td>
<td>IPNMAT</td>
<td>y</td>
<td>y</td>
<td></td>
<td>IP: nondurable materials</td>
</tr>
<tr>
<td>16</td>
<td>IPMANSICS</td>
<td>y</td>
<td>y</td>
<td></td>
<td>IP: manufacturing (SIC)</td>
</tr>
<tr>
<td>17</td>
<td>IPB51222S</td>
<td>y</td>
<td>y</td>
<td></td>
<td>IP: residential utilities</td>
</tr>
<tr>
<td>18</td>
<td>IPFUELS</td>
<td>y</td>
<td>y</td>
<td></td>
<td>IP: fuels</td>
</tr>
<tr>
<td>19</td>
<td>CUMFNS</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>capacity utilization: manufacturing</td>
</tr>
<tr>
<td>20</td>
<td>HWI</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>help wanted index</td>
</tr>
<tr>
<td>21</td>
<td>HWIURATIO</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>ratio of help wanted to unemployed</td>
</tr>
<tr>
<td>22</td>
<td>CLF16OV</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>civilian labor force</td>
</tr>
<tr>
<td>23</td>
<td>CE16OV</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>civilian employment</td>
</tr>
<tr>
<td>24</td>
<td>UNRATE</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>civilian unemployment rate</td>
</tr>
<tr>
<td>25</td>
<td>UEMPMEAN</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>average duration of unemployment (weeks)</td>
</tr>
<tr>
<td>26</td>
<td>UEMPLT5</td>
<td>y</td>
<td>y</td>
<td></td>
<td>civilian unemployed: less than 5 weeks</td>
</tr>
<tr>
<td>27</td>
<td>UEMP5TO14</td>
<td>y</td>
<td>y</td>
<td></td>
<td>civilian unemployed: 5-14 weeks</td>
</tr>
<tr>
<td>28</td>
<td>UEMP15OV</td>
<td>y</td>
<td>y</td>
<td></td>
<td>civilian unemployed: 15 weeks and over</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>UEMP15T26</td>
<td>y y</td>
<td>civilian unemployed: 15-26 weeks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>UEMP27OV</td>
<td>y y</td>
<td>civilian unemployed: 27 weeks and over</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>CLAIMSx</td>
<td>y y y</td>
<td>initial claims (for unemployment benefits)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>PAYEMS</td>
<td>y y y</td>
<td>all employees: total nonfarm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>USGOOD</td>
<td>y y</td>
<td>all employees: goods-producing industries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>CES1021000001</td>
<td>y y</td>
<td>all employees: mining</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>USCNS</td>
<td>y y</td>
<td>all employees: construction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>MANEMP</td>
<td>y y</td>
<td>all employees: manufacturing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>DMANEMP</td>
<td>y y</td>
<td>all employees: durable goods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>NDMANEMP</td>
<td>y y</td>
<td>all employees: non-durable goods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>SRVPRD</td>
<td>y y</td>
<td>all employees: service-providing industries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>USTPU</td>
<td>y y</td>
<td>all employees: trade, transportation and utilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>USWTRADE</td>
<td>y y</td>
<td>all employees: wholesale trade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>USTRADE</td>
<td>y y</td>
<td>all employees: retail trade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>USFIRE</td>
<td>y y</td>
<td>all employees: financial activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>USGOVT</td>
<td>y y</td>
<td>all employees: government</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>CES0600000007</td>
<td>y y y</td>
<td>average weekly hours: goods-producing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>AWOTMAN</td>
<td>y y</td>
<td>average weekly overtime hours: manufacturing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>AWHMAN</td>
<td>y y y</td>
<td>average weekly hours: manufacturing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>HOUST</td>
<td>y y y</td>
<td>housing starts: total new privately owned</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>HOUSTNE</td>
<td>y y</td>
<td>housing starts: northeast</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>HOUSTMW</td>
<td>y y</td>
<td>housing starts: midwest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>HOUSTS</td>
<td>y y</td>
<td>housing starts: south</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>HOUSTW</td>
<td>y y</td>
<td>housing starts: west</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>PERMIT</td>
<td>y y y</td>
<td>new private housing permits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>PERMITNE</td>
<td>y y</td>
<td>new private housing permits: northeast</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>PERMITMW</td>
<td>y y</td>
<td>new private housing permits: midwest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>PERMITS</td>
<td>y y</td>
<td>new private housing permits: south</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>PERMITW</td>
<td>y y</td>
<td>new private housing permits: west</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>AMDMNOx</td>
<td>y y y</td>
<td>new orders of durable goods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>ANDENOx</td>
<td>y y y</td>
<td>new orders for nondefense capital goods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>AMDMUOx</td>
<td>y y y</td>
<td>unfilled orders for durable goods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>BUSINVx</td>
<td>y y y</td>
<td>total business inventories</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>ISRATIOx</td>
<td>y y y</td>
<td>total business: inventories to sales ratio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>M1SL</td>
<td>y y y</td>
<td>M1 money stock</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>M2SL</td>
<td>y y y</td>
<td>M2 money stock</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>M2REAL</td>
<td>y y y</td>
<td>real M2 money stock</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>AMBSL</td>
<td>y y y</td>
<td>St Louise adjusted monetary base</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>TOTRESNS</td>
<td>y y y</td>
<td>total reserves of depository institutions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>NONBORRES</td>
<td>y y y</td>
<td>reserves of depository institutions, nonborrowed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>BUSLOANS</td>
<td>y y y</td>
<td>commercial and industrial loans</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>REALLN</td>
<td>y y y</td>
<td>real estate loans at all commercial banks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>NONREVSLS</td>
<td>y y y</td>
<td>total nonrevolving credit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>CONSPI</td>
<td>y y y</td>
<td>nonrevolving consumer credit to personal income</td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>S&amp;P 500</td>
<td>y y y</td>
<td>S&amp;P’s common stock price index: composite</td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>S&amp;P: industrials</td>
<td>y y y</td>
<td>S&amp;P’s common stock price index: industrials</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>S&amp;P div yield</td>
<td>y y y</td>
<td>S&amp;P’s composite common stocks: dividend yield</td>
<td></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>S&amp;P PE ratio</td>
<td>y y y</td>
<td>S&amp;P’s composite common stocks: price-earning ratio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>FEDFUNDS</td>
<td>y y</td>
<td>effective federal funds rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>CP3Mx</td>
<td>y y</td>
<td>3-month AA financial commercial paper rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>TB3MS</td>
<td>y y</td>
<td>3-month treasury bill rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>TB6MS</td>
<td>y y</td>
<td>6-month treasury bill rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>GS1</td>
<td>y y</td>
<td>1-year treasury bill rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>GS5</td>
<td>y y</td>
<td>5-year treasury bill rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>GS10</td>
<td>y y</td>
<td>10-year treasury bill rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>AAA</td>
<td>y y</td>
<td>Moody’s seasoned Aaa Corporate bond yield</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>BAA</td>
<td>y y</td>
<td>Moody’s seasoned Baa Corporate bond yield</td>
<td></td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>COMPAPFFFx</td>
<td>y</td>
<td>3 month commercial paper minus federal funds rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>TB3SMFFFM</td>
<td>y y</td>
<td>3 month treasury bill minus federal funds rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>TB6SMFFFM</td>
<td>y y</td>
<td>6 month treasury bill minus federal funds rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>T1YFFFM</td>
<td>y y</td>
<td>1-year treasury bill minus federal funds rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>T5YFFFM</td>
<td>y y</td>
<td>5-year treasury bill minus federal funds rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>T10YFFFM</td>
<td>y y</td>
<td>10-year treasury bill minus federal funds rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>AAAFFFM</td>
<td>y y</td>
<td>Moody’s Aaa corporate bond minus federal funds rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>BAAFFFM</td>
<td>y y</td>
<td>Moody’s Baa corporate bond minus federal funds rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>TWEXMMTH</td>
<td>y y y</td>
<td>trade weighted US dollar index: major currencies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>EXSZUSx</td>
<td>y y y</td>
<td>Switzerland/US foreign exchange rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>EXJPUSx</td>
<td>y y y</td>
<td>Japan/US foreign exchange rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>EXUSUKx</td>
<td>y y y</td>
<td>US/UK foreign exchange rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>EXCAUSx</td>
<td>y y y</td>
<td>Canada/US foreign exchange rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>WPSFD49207</td>
<td>y y y</td>
<td>PPI: finished goods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>WPSFD49502</td>
<td>y y y</td>
<td>PPI: finished consumer goods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>WPSID61</td>
<td>y y y</td>
<td>PPI: intermediate materials</td>
<td></td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>WPSID62</td>
<td>y y y</td>
<td>PPI: crude materials</td>
<td></td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>OILPRICEx</td>
<td>y y y</td>
<td>crude oil, spliced WTI and cushing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>PPICMMM</td>
<td>y y y</td>
<td>PPI: metals and metal products</td>
<td></td>
<td></td>
</tr>
<tr>
<td>105</td>
<td>CPIAUCSL</td>
<td>y y y</td>
<td>CPI: all items</td>
<td></td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>CPIAPPSSL</td>
<td>y y</td>
<td>CPI: apparel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>CPITRNSL</td>
<td>y y</td>
<td>CPI: transportation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>CPIMEDSL</td>
<td>y y</td>
<td>CPI: medical care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>CUSR0000SAC</td>
<td>y y</td>
<td>CPI: commodities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>CUUR0000SAD</td>
<td>y y</td>
<td>CPI: durables</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Code</td>
<td>y y y</td>
<td>Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------------------</td>
<td>-------</td>
<td>--------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>CUSR0000SAS</td>
<td></td>
<td>CPI: services</td>
<td></td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>CPIULFSL</td>
<td></td>
<td>CPI: all items less food</td>
<td></td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>CUUR0000SA0L2</td>
<td></td>
<td>CPI: all items less shelter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>114</td>
<td>CUSR0000SA0L5</td>
<td></td>
<td>CPI: all items less medical care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>PCEPI</td>
<td></td>
<td>personal consumption expenditure: chain index</td>
<td></td>
<td></td>
</tr>
<tr>
<td>116</td>
<td>DDURRG3M086SBEA</td>
<td>y y y</td>
<td>personal consumption expenditure: durable goods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>117</td>
<td>DNDGRG3M086SBEA</td>
<td>y y y</td>
<td>personal consumption expenditure: non durable goods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>DSERRG3M086SBEA</td>
<td>y y y</td>
<td>personal consumption expenditure: service</td>
<td></td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>CES0600000008</td>
<td>y y y</td>
<td>average hourly earnings: good producing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>CES2000000008</td>
<td>y y y</td>
<td>average hourly earnings: construction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>CES3000000008</td>
<td>y y y</td>
<td>average hourly earnings: manufacturing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>122</td>
<td>UMCSENTx</td>
<td>y y y</td>
<td>consumer sentiment index</td>
<td></td>
<td></td>
</tr>
<tr>
<td>123</td>
<td>MZMSL</td>
<td>y y y</td>
<td>MZM money stock</td>
<td></td>
<td></td>
</tr>
<tr>
<td>124</td>
<td>DTCOLNVHFNM</td>
<td>y y y</td>
<td>consumer motor vehicle loans outstanding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>DTCTHFNM</td>
<td>y y y</td>
<td>total consumer loans and leases outstanding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>126</td>
<td>INVEST</td>
<td>y y y</td>
<td>securities in bank credit at all commercial banks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>127</td>
<td>VXOCLSx</td>
<td>y y y</td>
<td>VXO</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Appendix C  Selected Macro and Finance Variables for Factors

Table 4: Variables Selected by Adaptive LASSO (AIC)

<table>
<thead>
<tr>
<th>FAR Factors</th>
<th>FPCA factors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Set A</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Factor 1</strong></td>
<td>Factor 1</td>
</tr>
<tr>
<td>10-year treasury bill rate</td>
<td>10-year treasury bill rate</td>
</tr>
<tr>
<td><strong>Factor 2</strong></td>
<td>Factor 2</td>
</tr>
<tr>
<td>none</td>
<td>3-month treasury bill rate</td>
</tr>
<tr>
<td><strong>Factor 3</strong></td>
<td>Factor 3</td>
</tr>
<tr>
<td>all employees: wholesale trade</td>
<td>all employees: mining</td>
</tr>
<tr>
<td>new private housing permits</td>
<td>all employees: government</td>
</tr>
<tr>
<td>federal funds rate</td>
<td>average weekly hours: goods-producing</td>
</tr>
<tr>
<td>1-year treasury bill minus federal funds rate</td>
<td>housing starts: total new privately owned</td>
</tr>
<tr>
<td>Canada/US foreign exchange rate</td>
<td>new private housing permits</td>
</tr>
<tr>
<td><strong>Factor 4</strong></td>
<td>Factor 4</td>
</tr>
<tr>
<td>IP: durable consumer goods</td>
<td>1-year treasury bill minus federal funds rate</td>
</tr>
<tr>
<td>all employees: mining</td>
<td>5-year treasury bill rate</td>
</tr>
<tr>
<td>all employees: construction</td>
<td>5-year treasury bill minus federal funds rate</td>
</tr>
<tr>
<td>all employees: financial activities</td>
<td>trade weighted US dollar index: major currencies</td>
</tr>
<tr>
<td>average weekly hours: goods-producing</td>
<td>Japan/US foreign exchange rate</td>
</tr>
<tr>
<td>new private housing permits</td>
<td>volatility index</td>
</tr>
<tr>
<td>S&amp;ps composite common stocks: price-earning ratio</td>
<td>none</td>
</tr>
<tr>
<td>5-year treasury bill rate</td>
<td></td>
</tr>
<tr>
<td>3 month commercial paper minus federal funds rate</td>
<td></td>
</tr>
<tr>
<td>1-year treasury bill minus federal funds rate</td>
<td></td>
</tr>
<tr>
<td>5-year treasury bill minus federal funds rate</td>
<td></td>
</tr>
<tr>
<td>volatility index</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Set B</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Factor 1</strong></td>
</tr>
<tr>
<td>none</td>
</tr>
<tr>
<td><strong>Factor 2</strong></td>
</tr>
<tr>
<td>none</td>
</tr>
<tr>
<td><strong>Factor 3</strong></td>
</tr>
<tr>
<td>IP: business equipment</td>
</tr>
</tbody>
</table>
Table 5: Variables Selected by Adaptive LASSO (BIC)

<table>
<thead>
<tr>
<th>FAR Factors</th>
<th>FPCA Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Set A</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Factor 1</strong></td>
<td><strong>Factor 1</strong></td>
</tr>
<tr>
<td>10-year treasury bill rate</td>
<td>10-year treasury bill rate</td>
</tr>
<tr>
<td><strong>Factor 2</strong></td>
<td><strong>Factor 2</strong></td>
</tr>
<tr>
<td>none</td>
<td>3-month treasury bill rate</td>
</tr>
<tr>
<td><strong>Factor 3</strong></td>
<td><strong>Factor 3</strong></td>
</tr>
<tr>
<td>all employees: wholesale trade</td>
<td>all employees: mining</td>
</tr>
<tr>
<td>new private housing permits</td>
<td>average weekly hours: goods-producing</td>
</tr>
<tr>
<td>Canada/US foreign exchange rate</td>
<td>housing starts: total new privately</td>
</tr>
<tr>
<td></td>
<td>owned</td>
</tr>
<tr>
<td></td>
<td>new private housing permits</td>
</tr>
<tr>
<td></td>
<td>5-year treasury bill rate</td>
</tr>
<tr>
<td><strong>Factor 4</strong></td>
<td></td>
</tr>
<tr>
<td>all employees: construction</td>
<td></td>
</tr>
<tr>
<td>new private housing permits</td>
<td></td>
</tr>
<tr>
<td>1-year treasury bill rate</td>
<td></td>
</tr>
<tr>
<td>volatility index</td>
<td></td>
</tr>
</tbody>
</table>
all employees: mining 1-year treasury bill minus federal funds rate
all employees: construction 5-year treasury bill minus federal funds rate
all employees: financial activities volatility index
new private housing permits
S&Ps composite common stocks: price-earning ratio
1-year treasury bill minus federal funds rate
5-year treasury bill minus federal funds rate
volatility index

<table>
<thead>
<tr>
<th>Factor 4</th>
<th>Factor 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>none</td>
</tr>
</tbody>
</table>

Set B

<table>
<thead>
<tr>
<th>Factor 1</th>
<th>Factor 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>none</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor 2</th>
<th>Factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>none</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor 3</th>
<th>Factor 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>all employees: wholesale trade</td>
<td>none</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor 4</th>
<th>Factor 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>new private housing permits</td>
<td>none</td>
</tr>
</tbody>
</table>

Set C

<table>
<thead>
<tr>
<th>Factor 1</th>
<th>Factor 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-year treasury bill rate</td>
<td>10-year treasury bill rate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor 2</th>
<th>Factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>3-month treasury bill rate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor 3</th>
<th>Factor 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>all employees: wholesale trade</td>
<td>none</td>
</tr>
<tr>
<td>new private housing permits</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factor 4</th>
<th>Factor 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>all employees: construction</td>
<td>none</td>
</tr>
<tr>
<td>new private housing permits</td>
<td></td>
</tr>
<tr>
<td>1-year treasury bill rate</td>
<td></td>
</tr>
<tr>
<td>volatility index</td>
<td></td>
</tr>
</tbody>
</table>
References


