Asymptotics of Functional Spectral Component Analysis with Weakly Dependent Data

Bo Hu

Institute of New Structural Economics

Peking University

AMES 2023
Singapore

July 30, 2023

Functional Data Analysis

From scalar, vector variables to functional variables

- Each observation is viewed as a realization of a random curve
- Functional data analysis

Scenarios

- Intergenerational mobility
- Asset return distributional dynamics
- Yield curve dynamics

Appl. 1: Intergenerational Mobility

Intergenerational Elasticity (IGE) Estimation:

$$\log Y_c = \alpha + \beta \log Y_p + \varepsilon$$

- Usually Y is the permanent income
- Timing of parental income also matters (early childhood vs later childhood, Carneiro et al. 2021)
- Can we just put throw all the yearly parental income in the regression?

$$\log Y_c = \alpha + \sum_{t=1}^{20} \beta_t \log Y_{pt} + \varepsilon$$

Appl. 1: Intergenerational Mobility

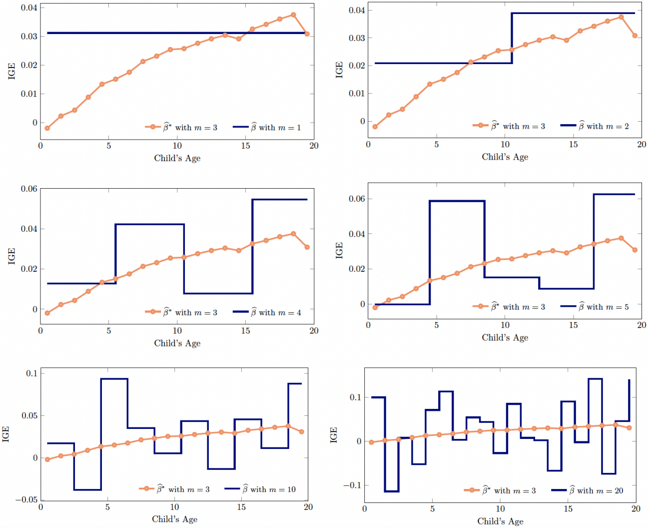

Figure: Chang et al. (2023)

Appl. 1: Intergenerational Mobility

Consider the (continuous) parental income trajectory $Y_p(t)$ as a functional variable

$$\log y_c = \int_0^{20} \beta(t) y_p(t) \mathrm{d}t + \varepsilon_t$$

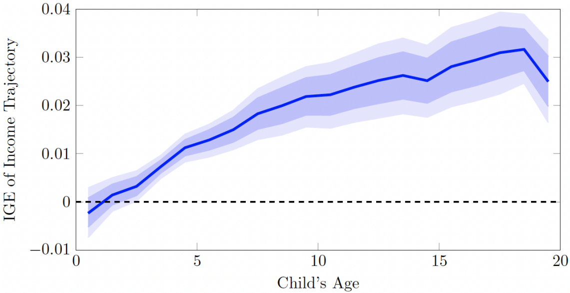

Figure: Chang el al. (2023)

Appl. 2: Asset Return Distributional Dynamics

How does distribution of asset return evolve across time?

- ARMA-class: models for conditional mean
  - $\mathbb{E}(r_t|\mathcal{F}_{t-1}) = \alpha_0 + \alpha_1 r_{t-1}$
- ARCH-class: models for conditional variance
  - $\mathsf{Var}(r_t|\mathcal{F}_{t-1}) = \alpha_0 + \alpha_1 r_{t-1}^2$
- VaR-class: models for tail probabilities
- Issues:
  - Model specific features of distribution
  - Assume specific dependence structure
  - No intrinsic guarantee of compatibility across models
- Why not look at distributions directly?
  - View observed distribution as realization of some random distribution
  - Distribution $\rightarrow$ density curves f_t
  - Model: $f_t = T(f_{t-1}) + \varepsilon_t$

Appl. 2: Asset Return Distributional Dynamics

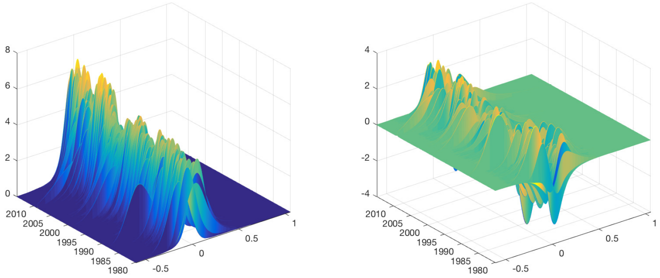

Figure: Densities and Demeaned Densities of the NYSE Stocks Monthly Returns

Appl. 2: Asset Return Distributional Dynamics

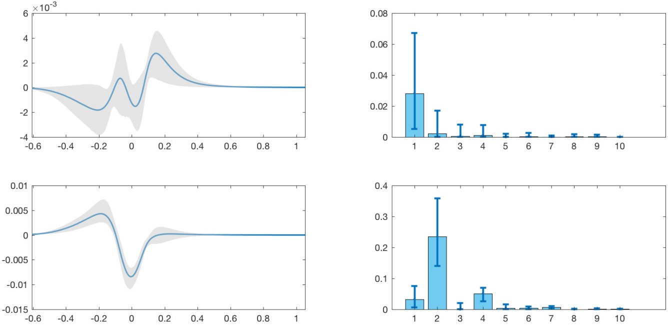

Figure: The Response Functions and the Forecast Variance Decompositions of the First Two Moments for the Density Process of the NYSE Stocks Monthly Returns

Appl. 2: Asset Return Distributional Dynamics

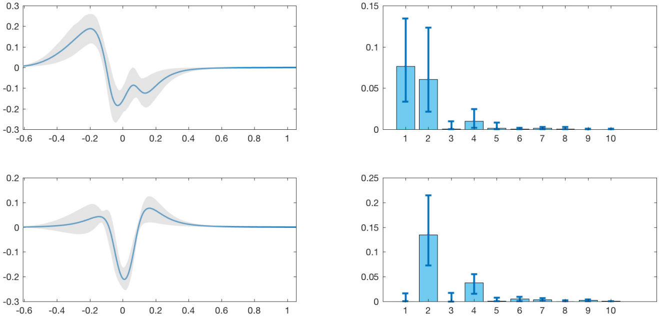

Figure: The Response Functions and the Forecast Variance Decompositions of the Tail Probabilities for the Density Process of the NYSE Stocks Monthly Returns

Appl. 3: Yield Curve Dynamics

Yield curve is the plot of bond yield against bond maturity (term structure of interest rate)

- Contains important information about financial mkt and macroeconomy
- Changes across time
- Time series models of yield at a single maturity (e.g., AR):
  - Ignore correlations of yields across different maturities
  - Ignore information in shape
- Time series models of yields at multiple maturities (e.g., VAR/ECM):
  - Alignment problem: bonds exists at particular maturity
  - 1-month, 2-month, 6-month, 5-year bonds becomes 1-month, 5-month, 11-month, $4\frac{11}{12}$-year bonds in the next month
- Estimate yield curve from bond yields, and model yield curve directly

Appl. 3: Yield Curve Dynamics

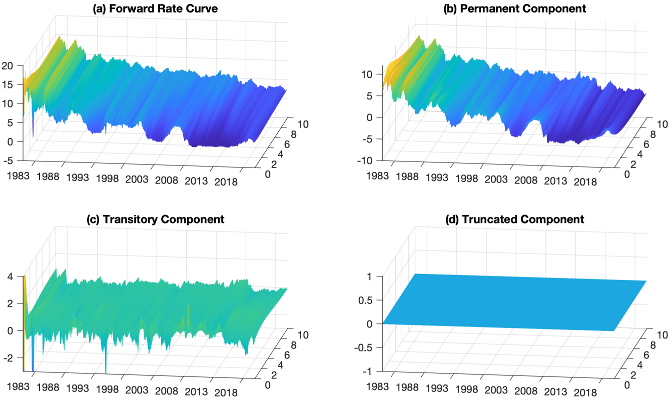

Figure: Time Series of Forward Rate Curves and Its Decompositions

Appl. 3: Yield Curve Dynamics

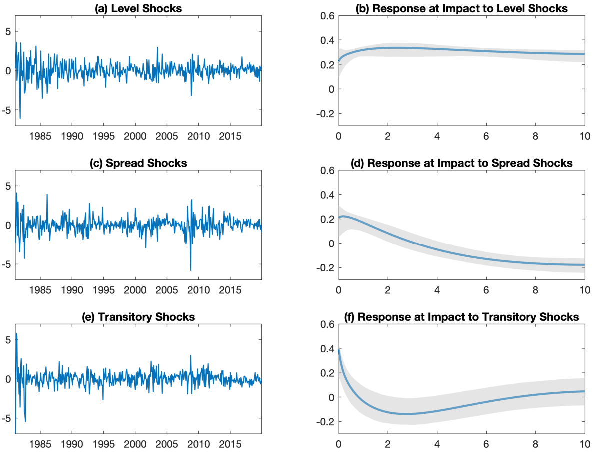

Figure: Structural Forward Rate Shocks

Appl. 3: Yield Curve Dynamics

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

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

Functional Spectral Component Analysis

FSCA is widely used in functional data analysis

- ◇ analysis based on spectrum of operators of functional data
- ◇ most significant example is FPCA
- ◇ important for understanding variance/covariance in the sample
- ◇ useful (and optimal) decomposition/dimension reduction tool
  - ◦ factor analysis
  - ◦ functional regressions
- ◇ both independent scenario and dependent scenario

It is necessary to understand asymptotic properties of spectrum related statistical quantities

- ◇ iid case: full result for FPCA
  - ◦ Dauxois et al. (1982)
- ◇ non-iid case: specific problems/partial results
  - ◦ Bosq (2000), Mas (2007), Hörmann and Kokoszka (2010), Hu et al. (2021)

Contributions

In this work, we provide

- asymptotic distribution theory for quantities related to FSCA in weakly dependent data setting in a unified approach
  - eigen-elements in FPCA
  - regularized estimators in ill-posed inverse problems
  - singular value decomposition for non-self adjoint operators
  - spectral decomposition for non-self adjoint operators
  - express these quantities as functions of spectrum of appropriate operators
  - use functional delta method to obtain asymptotic distribution
- CLTs for the second moment quantities of weakly dependent processes
- representations for one-dimensional projections of the above quantities so that they can be easily implemented in practice
- a procedure to determine the truncation parameter in some FPCA problems

Preliminaries: Hilbert-Valued Random Elements

The probability space $(\Omega, \mathcal{F}, \mathbb{P})$

H: a real separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$

A Borel measurable mapping $\xi : \Omega \to H$ is called an H-valued random element

$\xi \in L^p(H)$ if $\mathbb{E}\,\|\xi\|^p < \infty$

$\mathbb{E}\xi$ is defined as an element in H such that $\langle \mathbb{E}\xi, v\rangle = \mathbb{E}\langle \xi, v\rangle$ holds for any $v \in H$

- $\mathbb{E}\xi$ exists if $\xi \in L^1(H)$
- $\mathbb{E}$ is a linear and continuous operator

Preliminaries: Covariances

Covariance of ξ and η (assume both are mean zero) is defined by $\mathbb{E}(\xi \otimes \eta)$ where $x \otimes y$ may be viewed as

- a bilinear map $H \times H \to \mathbb{R}$ such that $(x \otimes y)(v_1, v_2) = \langle x, v_1 \rangle \langle y, v_2 \rangle$ for any $v_1, v_2 \in H$
- a linear map $H \to H$ such that $(x \otimes y)v = \langle x, v \rangle y$ for any $v \in H$

The tensor product $H_1 \otimes H_2$ is defined as the completion of the vector space spanned by $x \otimes y$ for $x \in H_1, y \in H_2$

- inner product in $H_1 \otimes H_2$: $\langle x_1 \otimes y_1, x_2 \otimes y_2 \rangle = \langle x_1, x_2 \rangle \langle y_1, y_2 \rangle$
- the inner product makes $H_1 \otimes H_2$ a separable Hilbert space
- $\mathbb{E}(\xi \otimes \eta)$ is therefore well defined
- $H_1 \otimes H_2$ can be identified with the space $L_{HS}(H_1, H_2)$ of all Hilbert-Schmidt operators from H_1 to H_2

Preliminaries: Algebra of Tensors

The definition of tensor product can be extended to Banach spaces with inner products replaced by linear functionals.

- $C \otimes D$ is well defined for C, D as bounded linear operators between Banach spaces
- $C \otimes D$ can be viewed as linear maps on $B_1 \otimes B_2$

Algebraic Properties

- $(x \otimes y)$ is linear
- $(x \otimes y)(x^*, y^*) = x^*(x)y^*(y)$
- $(x \otimes y)(x^*) = x^*(x)y$
- $(x \otimes y)^* = (y \otimes x)$
- $(C \otimes D)(x \otimes y) = (Cx) \otimes (Dy)$
- $(C \otimes D)[(x \otimes y) \otimes (z \otimes u)](E \otimes F) =$
  $[(E^*x) \otimes (Cy)] \otimes [(F^*z) \otimes (Du)]$

Preliminaries: Spectrum

Let $A \in L(H)$, the spectrum
$\sigma(A) = \{\lambda \in \mathbb{F} : \lambda I - A \text{ not invertible}\}$

An eigenvalue of A is $\lambda \in \mathbb{F}$ such that $\lambda I - A$ is not one-to-one.
$\mathcal{N}(\lambda I - A)$ is the eigenspace corresponding to the eigenvector λ

In spectral analysis

- Take $\mathbb{F} = \mathbb{C}$
- Consider the complexification $\mathbb{H}$ of H
- View (extend) operator A as operator on $\mathbb{H}$
- This does not affect our results since in the end all the spectral quantities (eigenvalues and singular values we encounter are real)

Preliminaries: Compact and Self-Adjoint Operators

$A \in L(H)$ is compact if it is the limit of a sequence of finite rank operators

- $\sigma(A)$ is at most countable, the only possible limit point is 0
- SVD: $A = \sum_{i=1}^{\infty} \mu_i (u_i \otimes w_i)$
  - μ_i are real, non-negative, can be arranged in descending order
  - μ_i^2 are eigenvalues of A^*A
  - u_i are eigenvectors of A^*A, and w_i are eigenvectors of AA^*

$A \in L(H)$ compact is self-adjoint if $A^* = A$

- We may arrange so that $\lambda_1 \geq \lambda_2 \geq \cdots \to 0$
- Spectral representation $A = \sum_{i=1}^{\infty} \lambda_i (v_i \otimes v_i)$
  - (λ, v_i) are eigen-pairs

Preliminaries: Functional Calculus

$A \in L(H)$
D: open set in $\mathbb{C}$ that includes $\sigma(A)$
f: holomorphic function on D
Γ: contour surrounds $\sigma(A)$ in D

Can define

$$f(A) = \frac{1}{2\pi i} \oint_\Gamma f(z)(zI - A)^{-1} \mathrm{d}z$$

(in the Riemann-Stieltjes integral sense.) The definition of $f(A)$ is independent of choices of Γ

In particular

- $A^n = \frac{1}{2\pi i} \oint_\Gamma z^n (zI - A)^{-1} \mathrm{d}z$
- Split $\sigma(A), D$ such that $\sigma_1 \subset D_1, \sigma_2 \subset D_2$, and take $f(z) = 1_{D_1}$, the $f(A)$ is the projection P onto the eigenspace $E(\sigma_1)$ along the direction of the eigenspace $E(\sigma_2)$

Functional Delta Method

Frechet derivative of $f(A) : L(H) \to L(H)$:

$$f'(A)\Pi = \frac{1}{2\pi i} \oint_\Gamma f(z)(zI - A)^{-1}\Pi(zI - A)^{-1}\mathrm{d}z$$

Theorem 1

Let A, D, f, Γ be defined as above. Suppose that there are a normalizing sequence r_T and an estimator $\hat{A}_T$ of A such that $r_T(\hat{A}_T - A) \to_d \Xi$. Let f_T be a sequence of holomorphic functions on D such that for some $\delta > 0$,

$$\sup_{\{z:\rho(z,\sigma(A))<\delta\}} |f_T(z) - f(z)| = o_p(r_T^{-1}),$$

then

$$r_T(f_T(\hat{A}_T) - f(A)) \to_d \frac{1}{2\pi i} \oint_\Gamma f(z)(zI - A)^{-1}\Xi(zI - A)^{-1}\mathrm{d}z.$$

Functional Delta Method

- The Cauchy formula may be applied to simplify the limit distribution term
- This result applies to general bounded linear operators in $L(H)$. In many of the applications A is compact and self-adjoint (such as in the case of FPCA), and one can utilize the resolution of the identity to rewrite the contour integral on the right hand side
- In many applications we only need results for the case when $f_T = f$ for all T. Here we provide a more general result.
- Many weak convergence problems in functional setting, in particular, problems related to FPCA, can be dealt with in this unified approach as long as we can find the proper A and f_T

I. Functional Principal Component Analysis

$\{X_t\}$: a strictly stationary sequence of H-valued random elements.
Assume $\mathbb{E}X_t = 0$ at the moment

FPCA based on the spectral representation of its variance

$$V = \mathbb{E}(X_t \otimes X_t) = \sum_{i=1}^{\infty} \lambda_i (v_i \otimes v_i)$$

The empirical counterpart is

$$\hat{V} = \frac{1}{T} \sum_{t=1}^{T} (X_t \otimes X_t) = \sum_{i=1}^{\infty} \hat{\lambda}_i (\hat{v}_i \otimes \hat{v}_i)$$

All quantities with hats are also dependent on the sample size T.
In asymptotics we let $T \to \infty$

We are interested in the asymptotic distributions of $\hat{\lambda}_i, \hat{v}_i$, and $\hat{P}_i = \hat{v}_i \otimes \hat{v}_i$

I. Functional Principal Component Analysis

Assumption 1

$\mathcal{N}(V) = \{0\}$, and V has no repeated eigenvalues so that we order the eigenvalues as $\lambda_1 > \lambda_2 > \cdots$

Assumption 2

$\sqrt{T}(\hat{V} - V) \to_d \mathbb{N}(0, K)$ for some $K \in (H \otimes H) \otimes (H \otimes H)$.

To obtain the asymptotic distributions of eigen-elements corresponidng to λ_i, We split $\sigma(V)$ and D so that D_1 contains λ_i and D_2 contains the rest eigenvalues. We take $f = 1_{D_1}$. Then $P_i = f(V)$, and λ_i and v_i can be written as linear functions of P_i.

I. Functional Principal Component Analysis

Theorem 2

Under Assumptions 1 and 2,

$$\sqrt{T}(\hat{\lambda}_i - \lambda_i) \to_d \langle U v_i, v_i \rangle,$$

$$\sqrt{T}(\hat{v}_i - v_i) \to_d Q_i U v_i,$$

and

$$\sqrt{T}\left(\hat{P}_i - P_i\right) \to_d P_i U Q_i + Q_i U P_i$$

where U is an $\mathbb{N}(0, K)$ random element and $Q_i = \sum_{k \neq i} \frac{1}{\lambda_i - \lambda_k} P_k$. The convergences also hold jointly. Note that all the limit distributions are Gaussian.

The analysis could also be applied to the PCA of the long run variance operator of X_t.

I. Functional Principal Component Analysis

Sometimes we are interested in the orthogonal projection $\Pi_K = \sum_{i=1}^{K} P_i$ onto the eigenspace corresponding to the first K eigenvalues. Split D into $D_1 \cup D_2$ so that D_1 contains the first K eigenvalues, and take $f(z) = 1_{D_1}(z)$, we have

Theorem 3

Under Assumptions 1 and 2,

$$\sqrt{T}\left(\hat{\Pi}_K - \Pi_K\right) \to_d \sum_{i=1}^{K} (P_i U Q_i + Q_i U P_i)$$

where U is an $\mathbb{N}(0, K)$ random element and $Q_i = \sum_{k=K+1} \frac{1}{\lambda_i - \lambda_k} P_k$.

- If we allow K to change with sample size, we have that $\left\|\hat{\Pi}_K - \Pi_K\right\| = O_p\left(\frac{1}{\sqrt{T}} \sum_{i=1}^{K} \frac{1}{\lambda_i - \lambda_{K+1}}\right)$

I. Functional Principal Component Analysis

We may also be interested in the inverse of V. Since V is not invertiable, we may consider the pseudo inverse with Tikhnov regularization: $V^\dagger = (V + \alpha I)^{-1}$ for some $\alpha \neq 0$. Taking $f(z) = (z + \alpha)^{-1}$ we have

Theorem 4

Under Assumptions 1 and 2,

$$\sqrt{T}\left(\hat{V}^\dagger - V^\dagger\right) \to_d \mathcal{S}U\mathcal{S}$$

where U is an $\mathbb{N}(0, K)$ random element and $\mathcal{S} = \sum_{i=1}^{\infty} \frac{1}{\lambda_i + \alpha} P_i$

- If we allow α to change with sample size, we have that $\left\|\hat{V}^\dagger - V^\dagger\right\| = O_p(\frac{1}{a_T^2 \sqrt{T}})$.

I. Functional Principal Component Analysis

Another approach is to consider the pseudo inverse on particular subspaces: $V^{+} = \sum_{i=1}^{K} \lambda_i^{-1} (v_i \otimes v_i)$. Split D into $D_1 \cup D_2$ so that D_1 contains the first K eigenvalues, and take $f(z) = 1_{D_1}(z) z^{-1}$, we have

Theorem 5

Under Assumptions 1 and 2,

$$\sqrt{T}\left(\hat{V}^{+} - V^{+}\right) \to_d \sum_{i=1}^{K} (\mathcal{P}_i U \mathcal{Q}_i + \mathcal{Q}_i U \mathcal{P}_i - \mathcal{P}_i U \mathcal{P}_i)$$

where U is an $\mathbb{N}(0, K)$ random element $\mathcal{P}_i = \frac{1}{\lambda_i} P_i$, and $\mathcal{Q}_i = \sum_{j \neq i} \frac{1}{\lambda_j - \lambda_i} P_i$.

- If we allow K to change with sample size, we have that $\left\| \hat{V}^{+} - V^{+} \right\| = O_p\left(\frac{1}{\sqrt{T}} \sum_{i=1}^{K} \frac{1}{\lambda_i} \sum_{j \neq i} \frac{1}{\lambda_i - \lambda_j} \right)$

II. Singular Value Decomposition Analysis

Let A be a compact operator in $L(H)$ and $\hat{A}$ its estimator

We have singular value decompositions

$$A = \sum_{i=1}^{\infty} \mu_i (u_i \otimes w_i), \qquad \hat{A} = \sum_{i=1}^{\infty} \hat{\mu}_i (\hat{u}_i \otimes \hat{w}_i)$$

This decomposition may be used to analyze, for example, the magnitude of the autocovariances of $\{X_t\}$ of all orders.

Assumption 3

$\mathcal{N}(A) = \mathcal{N}(A^*) = \{0\}$, and A has no repeated singular values so that we order the singular values $\mu_1 > \mu_2 > \cdots$.

Assumption 4

$\sqrt{T}(\hat{A} - A) \to_d \mathbb{N}(0, \mathcal{K})$ for some $\mathcal{K} \in (H \otimes H) \otimes (H \otimes H)$.

II. Singular Value Decomposition Analysis

To obtain asymptotics, we utilize the relationship between the singular value decomposition of A and the eigen-decompositions of A^*A and AA^*.

Theorem 6

Under Assumptions 3 and 4,

$$\sqrt{T}(\hat{\mu}_i - \mu_i) \to_d \frac{1}{2\mu_i}\langle (A^*\mathcal{U} + \mathcal{U}^*A)u_i, u_i\rangle,$$

$$\sqrt{T}(\hat{u}_i - u_i) \to_d Q_{u_i}(A^*\mathcal{U} + \mathcal{U}^*A)u_i,$$

and

$$\sqrt{T}(\hat{w}_i - w_i) \to_d Q_{w_i}(A\mathcal{U}^* + \mathcal{U}A^*)w_i$$

where $\mathcal{U}$ *is an* $\mathbb{N}(0, \mathcal{K})$ *random element,* $Q_{u_i} = \sum_{k\neq i} \frac{u_k \otimes u_k}{\mu_i^2 - \mu_k^2}$, *and* $Q_{w_i} = \sum_{k\neq i} \frac{w_k \otimes w_k}{\mu_i^2 - \mu_k^2}$. *The convergences also hold jointly. Note that all the limit distributions are Gaussian.*

III. Spectral Decomposition Analysis

In multi-dimensional or high-dimensional setting

- properties of the series in different subspaces are different
- these subspaces are characterized by the generalized eigenspaces of some operator
- e.g., Beveridge-Nelson decomposition

Let A be a compact operator in $L(H)$ and $\hat{A}$ its estimator

Assumption 5

$\mathcal{N}(A) = \{0\}$, and A has eigenvalues (without repeatition) $\lambda_1, \lambda_2, \ldots$ with algebraic multiplicity $m_1, m_2, \ldots$.

III. Spectral Decomposition Analysis

Estimate the subspace corresponding to the first K eigenvalues by

$$\bigoplus_{i=1}^{K} \mathcal{N}\left((\lambda_i I - \hat{A})^{m_i}\right),$$

The corresponding (possibly non-orthogonal) projection

$$\hat{P} = \frac{1}{2\pi i} \oint_{\Gamma} 1_{D_1}(z)(zI - \hat{A})^{-1} \mathrm{d}z$$

Theorem 7

Under Assumptions 5 and 4,

$$\sqrt{T}(\hat{P} - P) \to_d \frac{1}{2\pi i} \oint_{\Gamma} 1_{D_1}(z)(zI - A)^{-1} \mathcal{U}(zI - A)^{-1} \mathrm{d}z$$

where $\mathcal{U}$ *is an* $\mathbb{N}(0, \mathcal{K})$ *random element. Note that the limit distribution is Gaussian.*

Weak Dependence Concepts

Let $\mathcal{F}_m^n = \sigma(X_t, m \le t \le n)$. The sequence $\{X_t\}$ is called

1. α-mixing if
   $\alpha(k) = \sup_n \sup_{A \in \mathcal{F}_{-\infty}^n, B \in \mathcal{F}_{n+k}^\infty} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \to 0$
   as $k \to \infty$
2. ϕ-mixing if
   $\phi(k) = \sup_n \sup_{A \in \mathcal{F}_{-\infty}^n, B \in \mathcal{F}_{n+k}^\infty, \mathbb{P}(A)>0} |\mathbb{P}(B|A) - \mathbb{P}(B)| \to 0$
   as $k \to \infty$
3. L^p-m-approximable if $X_t = f(\varepsilon_t, \varepsilon_{t-1}, \cdots)$ for some measurable f and iid $\{\varepsilon_t\}$, and for each t there is an independent copy $\{\varepsilon_i^{(t)}\}$ of $\{\varepsilon_i\}$ such that $X_t^{(m)}$ defined by
   $X_t^{(m)} = f(\varepsilon_t, \varepsilon_{t-1}, \cdots, \varepsilon_{t-m+1}, \varepsilon_{t-m}^{(t)}, \varepsilon_{t-m-1}^{(t)}, \cdots)$ satisfies
   $\sum_{m=1}^\infty \left(\mathbb{E}\left\|X_t - X_t^{(t)}\right\|^p\right)^{1/p} < \infty$

To establish CLT, other weak dependence concepts for H-valued random elements may also be utilized

Central Limit Theorem for Sample Variance

Assumption 6

Suppose that one of the following conditions hold.

1. $\{X_t\}$ is an α-mixing strictly stationary sequence such that $\mathbb{E}\left\|X_t\right\|^{4+2\delta} < \infty$, and its α-mixing coefficients α_k satisfies $\sum_{k=1}^{\infty} \alpha_k^{\frac{\delta}{2+\delta}} < \infty$.
2. $\{X_t\}$ is a ϕ-mixing strictly stationary sequence such that $\mathbb{E}\left\|X_t\right\|^{4} < \infty$, and its ϕ-mixing coefficients $\phi(k)$ satisfies $\sum_{k=1}^{\infty} \phi_k^{\frac{1}{2}} < \infty$.
3. $\{X_t\}$ is a L^4-m-approximable sequence.

In the case when $\{X_t\}$ is not mean zero, we estimate V by

$$\hat{V} = \frac{1}{T} \sum_{t=1}^{T} \left[\left(X_t - \overline{X}_T\right) \otimes \left(X_t - \overline{X}_T\right) \right].$$

where $\overline{X}_T = \frac{1}{T} \sum_{t=1}^{T} X_t$.

Central Limit Theorem for Sample Variance

Theorem 8

Under Assumption 6,

$$\sqrt{T}(\hat{V} - V) \to_d \mathbb{N}\left(0, \sum_{h=-\infty}^{\infty} \kappa(h)\right)$$

where $\kappa : \mathbb{Z} \to (H \otimes H) \otimes (H \otimes H)$ *is defined by*

$$\begin{aligned}\kappa(h) =& \mathbb{E}[(X_h - \mathbb{E}X_h) \otimes (X_h - \mathbb{E}X_h) \otimes (X_0 - \mathbb{E}X_0) \otimes (X_0 - \mathbb{E}X_0)] \\ & - [\mathbb{E}((X_h - \mathbb{E}X_h) \otimes (X_h - \mathbb{E}X_h))] \otimes [\mathbb{E}((X_0 - \mathbb{E}X_0) \otimes (X_0 - \mathbb{E}X_0))] \,.\end{aligned}$$

Note that $\kappa(h)$ could be viewed as the autocovariance function of the $H \otimes H$-valued process $\{X_t \otimes X_t\}$.

Estimation of the Long Run Variance

To conduct statistical inferences using the above result, we need to estimate the long run variance operator $\sum_{h=-\infty}^{\infty} \kappa(h)$. The autocovariance operator could be estimated by

$$
\begin{aligned}
\hat{\kappa}(h) =& \frac{1}{T} \sum_{t=h+1}^{T} \Bigg[\Bigg((X_t - \overline{X}_T) \otimes (X_t - \overline{X}_T) - \frac{1}{T} \sum_{s=1}^{T} \left[(X_s - \overline{X}_T) \otimes (X_s - \overline{X}_T) \right] \Bigg) \\
& \otimes \Bigg((X_{t-h} - \overline{X}_T) \otimes (X_{t-h} - \overline{X}_T) - \frac{1}{T} \sum_{s=1}^{T} \left[(X_s - \overline{X}_T) \otimes (X_s - \overline{X}_T) \right] \Bigg) \Bigg]
\end{aligned}
$$

for $0 \leq h \leq T-1$, and $\hat{\kappa}(h) = \hat{\kappa}(-h)^*$ for $-(T-1) \leq h < 0$.

We then estimate the long run variance by

$$
\widehat{LRV}(X_t \otimes X_t) = \sum_{|h| \leq (T-1)} w(b_T h) \hat{\kappa}(h)
$$

where w is a suitable window function and b_T is the bandwidth parameter.

Estimation of the Long Run Variance

Assumption 7

1. $X_t \in L^8(H)$ and the fourth order cumulant $Q(r, s, t)$ of the process $\{X_t \otimes X_t\}$ is absolutely summable.
2. $w : \mathbb{R} \to \mathbb{R}_+$ is an even, bounded, square integrable function such that $w(0) = 1$ and that for every b and T we have $b \sum_{|h|<T} w(bh) \leq C(bT)^{1/2-\epsilon}$ for some $\epsilon > 0$
3. $\sum_{-\infty}^{\infty} |h|^q \|\kappa(h)\| < \infty$ for some $q > 0$.
4. There exists positive integer $r \geq q$ such that $\lim_{z\to 0} \frac{1-w(z)}{|z|^r} < \infty$ and is nonzero.
5. $b_T \to 0, b_T T \to \infty$, and $0 < \lim_{T\to\infty} b_T^{1+2q} T < \infty$.

Theorem 9

Under Assumptions 6 and 7, we have

$$\widehat{LRV}(X_t \otimes X_t) \to_p \sum_{h=-\infty}^{\infty} \kappa(h).$$

Representation

Sometimes it is useful to project the functional objects onto lower dimensional spaces (preferably finite dimensional spaces) and represent the limit distributions in more familiar forms.

We utilize the Karhunen-Loeve expansion for H-valued random elements:

$$X_t = \mathbb{E}X_t + \sum_{i=1}^{\infty} Z_{ti} v_i$$

where Z_{ti} is an array of real valued random variables such that $\mathbb{E}Z_{ti}^2 = \lambda_i$ and $\mathbb{E}Z_{ti}Z_{tj} = 0$ for $i \neq j$. Note that $\langle X_t - \mathbb{E}X_t, v_i \rangle = Z_{ti}$.

We next state a representation theorem corresponding to Theorem 5. Representation results for other theorems can be obtained similarly using algebraic rules of tensors introduced earlier.

Representation

Theorem 10

Under the assumptions of Theorem 5, we have the followings.

1. $\langle Uv_i, v_i \rangle =_d \mathbb{N}(0, LRV(Z_{ti}^2))$.
2. $\langle Q_i U v_i, v \rangle =_d \mathbb{N}\left(0, LRV\left(\sum_{j \neq i} \frac{Z_{ti} Z_{tj} \langle v_j, v \rangle}{\lambda_i - \lambda_j}\right)\right)$ *for any* $v \in H$*. In particular,* $\langle Q_i U v_i, v_j \rangle =_d \mathbb{N}\left(0, \frac{Z_{ti} Z_{tj}}{\lambda_i - \lambda_j}\right)$ *if* $j \neq i$*, and* $\langle Q_i U v_i, v_j \rangle$ *is degenerate if* $j = i$*.*
3. *For any* $u, v \in H$, $\big\langle (P_i U Q_i + Q_i U P_i) v, u \big\rangle =_d$ $\mathbb{N}\left(0, LRV\left(\sum_{j \neq i} \frac{Z_{ti} Z_{tj} \left(\langle v_i, v \rangle \langle v_j, u \rangle + \langle v_j, v \rangle \langle v_i, u \rangle\right)}{\lambda_i - \lambda_j}\right)\right)$*. In particular,* $\big\langle (P_i U Q_i + Q_i U P_i) v_j, v_k \big\rangle =_d \mathbb{N}\left(0, LRV\left(\frac{Z_{ti} Z_{tk}}{\lambda_i - \lambda_k}\right)\right)$ *if* $j = i, k \neq i$*,* $\big\langle (P_i U Q_i + Q_i U P_i) v_j, v_k \big\rangle =_d \mathbb{N}\left(0, LRV\left(\frac{Z_{ti} Z_{tj}}{\lambda_i - \lambda_j}\right)\right)$ *if* $j \neq i, k = i$*, and* $\big\langle (P_i U Q_i + Q_i U P_i) v_j, v_k \big\rangle$ *is degenerate for other combinations of* j *and* k*.*

The convergences also hold jointly.

Determining Truncation Parameters

As an application of our results, we propose a test to determine the truncation parameter in FPCA analysis

A frequently used criterion in selecting the truncation parameter is to choose K so that the first K principal components explain more than a θ proportion of total variation.

We therefore propose a sequence of one tailed test with null

$$H_0 : \frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{\infty} \lambda_i} = \theta$$

against the alternative

$$H_1 : \frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{\infty} \lambda_i} < \theta$$

Determining Truncation Parameters

The test is based on the statistic

$$\widetilde{T}_\theta(K) = \sqrt{T}\left(\frac{\sum_{i=1}^{K}\hat{\lambda}_i}{\sum_{i=1}^{\infty}\hat{\lambda}_i} - \theta\right).$$

Suppose Assumptions 1, 2 and 6 hold. Under the null we have

$$\widetilde{T}_\theta(K) \to_d \mathbb{N}\left(0, \frac{1}{(\sum_{i=1}^{\infty}\lambda)^2} LRV\left((1-\theta)\sum_{i=1}^{k} Z_{ti}^2 - \theta\sum_{i=k+1}^{\infty} Z_{ti}^2\right)\right)$$

where $Z_{ti} = \langle X_t - \mathbb{E}X_t, v_i\rangle$. A feasible version of the test is

$$T_\theta(K) = \frac{\sqrt{T}\left(\sum_{i\le k}\hat{\lambda}_i - \theta\sum_{i=1}^{\infty}\hat{\lambda}_i\right)}{\sqrt{\widehat{LRV}\left((1-\theta)\sum_{i=1}^{k}\hat{Z}_{ti}^2 - \theta\sum_{i=k+1}^{\infty}\hat{Z}_{ti}^2\right)}}$$

where $\widehat{LRV}$ is any consistent estimator of the long run variance. We have that under the null hypothesis, $T_\theta(K) \to_d \mathbb{N}(0,1)$.

Determining Truncation Parameters

The estimation of K is based on the sequential test (at a significance level of α) procedure as follows. Let $\Phi^{-1}(\alpha)$ be the α-quantile of the standard normal distribution.

1. Start from a large enough integer n
2. Conduct the test
   - If $T_\theta(n) > \Phi^{-1}(\alpha)$, we fail to reject null. We then replace n with $n-1$ and reconduct the test.
   - If $T_\theta(n) \leq \Phi^{-1}(\alpha)$, we reject the null, and stop.
3. Set $\hat{K} = n + 1$.

We have that under Assumptions 1, 2 and 6, $\hat{K}$ converges in probability to the true value.

Simulations

We simulate a strictly stationary series X_t where

- $X_t =_d \sum_{i=1}^{\infty} Z_{ti} v_i$
- $v_i(x) = \sqrt{2}\cos(i\pi x)$ defined on $[0, 1]$
- each Z_{ti}, in terms of t is an individual AR(1) process with autoregressive coefficient $1/2$ and variance i^{-3}
- the error term in the AR(1) processes are iid normal
- $\lambda_i = i^{-r}$, $\sum_{i=1}^{\infty} \lambda_i = \sum_{i=1}^{\infty} i^{-r} = \zeta(r)$ where $\zeta(\cdot)$ is the Riemann's zeta function.
- $\theta = 0.95$ which corresponds to $K = 3$.
- in estimating the long run variance we use the Bartlett kernel with Newey-West optimal bandwidth
- for each exercise we simulate 1000 samples

We use the test procedure above to select K. We try different combinations of sample size T and the number N of basis functions used in representing functions.

Simulations

	exact θ $K=3$	$\theta = 0.95$ $K=3$	$\theta = 0.95$ $K=2$
$T = 200$	0.062	0.000	0.661
$T = 300$	0.073	0.000	0.804
$T = 500$	0.052	0.000	0.923
$T = 1000$	0.049	0.000	0.998

Table: Rejection Ratio

Conclusions

Obtain asymptotic distributions of spectral-related quantities in weakly dependent data setting in a unified approach

- eigen-elements in FPCA
- regularized estimators in ill-posed inverse problems
- singular value decomposition for non-self adjoint operators
- spectral decomposition for non-self adjoint operators

Issues that can be explored using our results

- non-linear FPCA
- Inference of general FAR process
- Order selection of FMA models
- Optimal truncation parameter selection in FAR models
- ...