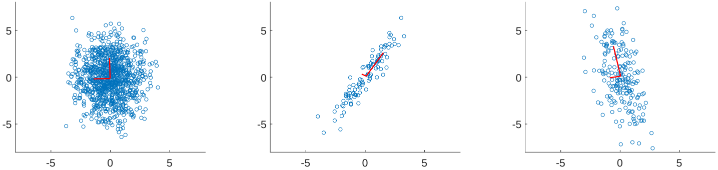
Algorithmic Trading: Functional Principal Component Analysis

Sebastian Jaimungal
University of Toronto

Jan, 2018

PCA

- **Principal component analysis** (PCA) aims to find the **independent** modes of fluctuations of data
- For example in 2-D



- PCAs are obtained by looking at the **eigenvectors** associated with the covariance matrix of the data
- **eigenvalues** correspond to the variance in the direction of that *mode of fluctuation*

PCA

- Let $\boldsymbol{\Sigma}$ denote the covariance matrix of $\mathbf{X}$
- Since $\boldsymbol{\Sigma}$ is positive-definite, there exists an **orthogonal matrix** U and a **diagonal matrix** D such that

$$\boldsymbol{\Sigma} = \boldsymbol{U}\,\mathbf{D}\,\boldsymbol{U}^{\intercal}$$

- Moreover, $\boldsymbol{U}$ corresponds to the matrix of **eigenvectors** of Σ (stacked columnwise) and the diagonal elements of $\mathbf{D}$ the corresponding **eigenvalues**.

$$\boldsymbol{\Sigma}\,\boldsymbol{U}_{\cdot k} = \mathbf{D}_{kk}\,\boldsymbol{U}_{\cdot k}$$

and $\boldsymbol{U}_{\cdot j}^{\intercal}\boldsymbol{U}_{\cdot k} = \delta_{jk}$

PCA

- Hence, writing $\mathbf{Y} = \boldsymbol{U}^\intercal \mathbf{X}$,

$$
\begin{aligned}
\mathbb{C}[\mathbf{Y}_i,\, \mathbf{Y}_j] &= \mathbb{C}\left[\sum_k \boldsymbol{U}_{ki}\,\mathbf{X}_k,\, \sum_l \boldsymbol{U}_{lj}\,\mathbf{X}_l\right] \\
&= \sum_{k,l} \boldsymbol{U}_{ki}\boldsymbol{U}_{lj}\mathbb{C}\left[\mathbf{X}_k,\, \mathbf{X}_l\right] \\
&= \sum_{k,l} \boldsymbol{U}_{ki}\boldsymbol{U}_{lj}\boldsymbol{\Sigma}kl \\
&= (\boldsymbol{U}^\intercal \boldsymbol{\Sigma} \boldsymbol{U})_{ij} \\
&= (\boldsymbol{U}^\intercal \boldsymbol{U} \mathbf{D} \boldsymbol{U}^\intercal \boldsymbol{U})_{ij} = \mathbf{D}_{ij}
\end{aligned}
$$

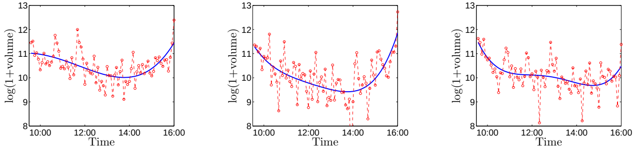
- The elements of the random variables $\mathbf{Y}$ are all **independent**, and have variance $\mathbf{D}_{ij}$
- They are the **principal components**

FPCA

- **Functional Principal Component Analysis (FPCA)** is the functional version of PCA
- The raw data $\{x_{i1}, \dots, x_{in_i}\}_{i=1,\dots,N}$ are viewed as noisy observations of a function at discrete points in "time" $\{t_{i1}, \dots, t_{in_i}\}_{i=1,\dots,N}$
- Take observation i, and **regress** the **observed values** $\{x_{ij}\}_{j=1,\dots,n_i}$ **onto a basis** $\phi(t) = \{\phi_1(t), \dots, \phi_K(t)\}$, i.e.

$$\boldsymbol{\alpha}_i = \arg\min_{\boldsymbol{\alpha}} \sum_{j=1}^{n_i} \left(x_{ij} - \sum_{k=1}^{K} \alpha_{i,k}\, \phi_k(t_{ij}) \right)^2$$

this can be done using **least-squares**.



(a) Oct-02-2014 (b) Oct-03-2014 (c) Oct-04-2014

FPCA

Many different basis choice... domain specific knowledge

- **Legendre polynomials** $P_n(t)$ basis, these are solutions to Legendre's ODE

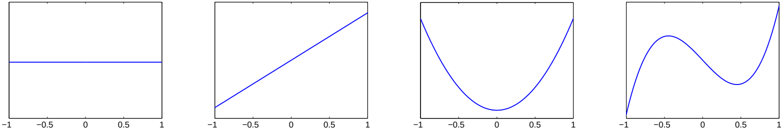
$$\frac{d}{dt}\left((1-t^2)\frac{d}{dt}P_n(t)\right)+n(n+1)P_n(t)=0$$

and can be written as

$$P_n(t)=\frac{1}{2^n\,n!}\frac{d^n}{dt^n}\left((t^2-1)^n\right)$$

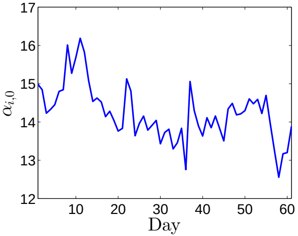
They form an **orthogonal basis** w.r.t. the L^2 inner product

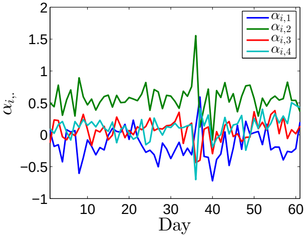
$$\int_{-1}^{1}P_m(t)\,P_n(t)\,dt=\frac{2}{2\,n+1}\delta_{mn}\,.$$



FPCA

This provides a "time-series" of coefficients $\boldsymbol{\alpha}_{i\cdot}$.





FPCA typically assumes that these are independent.. but there are modifications that account for functional time-series.
We will stick to the independent case.

FPCA

- Define the **covariance function**

$$K(s,t) = cov\left(x(s)\,,\,x(t)\right)$$

this is the analog of the **covariance matrix** $\boldsymbol{\Sigma}$ of a multi-variate sequence of observations

- Recall for PCA

$$\boldsymbol{\Sigma} = \boldsymbol{U}^\intercal\,\mathbf{D}\,\boldsymbol{U} = \sum_{j=1}^{M} d_j\,\boldsymbol{U}_{\cdot j}\,\boldsymbol{U}_{\cdot j}^\intercal \qquad \text{with} \qquad \boldsymbol{U}_{\cdot i}^\intercal\,\boldsymbol{U}_{\cdot j} = \delta_{ij}$$

fPCA

- **Mercer's lemma** states that if $K(s,t)$ is continuous on $\mathcal{D}^2$, then there exists an **orthonormal** sequence of continuous functions $\{\psi_i(t), i = 1, \dots\}$ (**eignefunctions**) such that

$$(\mathcal{K}\,\psi_i)(t) = \kappa_i\,\psi_i(t),$$

where the **kernel operator** $\mathcal{K}$ acts on square-integrable functions f as follows

$$(\mathcal{K}\,f)(t) = \int_{\mathcal{D}} K(s,t)\,f(s)\,ds\,.$$

Moreover, the covariance function

$$K(s,t) = \sum_{j=1}^{\infty} \kappa_j\,\psi_j(s)\,\psi_j(t)\,, \qquad \text{with} \qquad \int_{\mathcal{D}} \psi_i(t)\,\psi_j(t)\,dt = \delta_{ij}$$

and $\sum_{j=1}^{\infty} \kappa_j = \int_{\mathcal{D}} K(s,s)\,ds < +\infty$.

fPCA

- The **Karhunen-Loéve** expansion states that under the assumptions of Mercer's lemma, we have

$$x(t) = \mu(t) + \sum_{j=1}^{\infty} \sqrt{\kappa_j}\, \xi_j\, \psi_j(t)\,,$$

where,

$$\xi_j = \frac{1}{\sqrt{\kappa_j}} \int_{\mathcal{D}} x(s)\, \psi_j(s)\, ds$$

is a random variable with

$$\mathbb{E}[\xi_i] = 0 \qquad \text{and} \qquad \mathbb{E}[\xi_i\, \xi_j] = \delta_{ij}\,, \qquad \forall i, j \in \mathbb{N}\,,$$

and the series converges uniformly on $\mathcal{D}$ w.r.t the L^2 norm.

fPCA

- The eigenvalues κ_i can be interpreted as the size of the variation of $x(t)$ in the direction of $\psi_i(t)$
- We typically truncate the series to obtain a **dimensionally reduced approximation** of x

$$\hat{x} = \mu + \sum_{j=1}^{M} \sqrt{\kappa_j}\, \xi_k\, \psi_k(t)$$

- But **how to estimate** the eigenfunctions and eigenvalues from data?

fPCA

- Let the **sample mean function** be denoted $\overline{\mu}$, i.e.,

$$\overline{\mu}(t) = \tfrac{1}{n} \sum_{i=1}^{n} x_i(t)$$

- The empirical approximation of $K(s,t)$ is

$$\hat{K}(s,t) = \tfrac{1}{N} \sum_{i=1}^{N} \left(x_i(s) - \mu(s)\right)\left(x_i(t) - \mu(t)\right) .$$

- Mercer's lemma then implies there exists an orthonormal basis of eigenfunctions $\hat{\psi}_j$ with eigenvalues $\hat{\kappa}_j$ such that

$$\hat{K}(s,t) = \sum_{j=1}^{\infty} \hat{\kappa}_j \, \hat{\psi}_j(s) \, \hat{\psi}_j(t)$$

fPCA

- Expand the (centered) observations and eigenfunctions onto the basis functions

$$x_i(t) = \sum_{k=1}^{K} c_{ik}\phi_j(t), \qquad \text{and} \qquad \hat{\psi}_j(t) = \sum_{k=1}^{K} b_{jk}\phi_j(t)$$

- Then one can show that

$$\boldsymbol{b}_j = \boldsymbol{W}^{-\frac{1}{2}}\boldsymbol{u}_j$$

where $\boldsymbol{u}_j$ solves the eigenproblem

$$\boxed{\left(\tfrac{1}{N}\boldsymbol{W}^{\frac{1}{2}}\,\boldsymbol{C}^\intercal\,\boldsymbol{C}\,\boldsymbol{W}^{\frac{1}{2}}\right)\boldsymbol{u}_j = \kappa_j\,\boldsymbol{u}_j}$$

and

$$W_{kl} = < \phi_k, \phi_l > \ .$$

fPCA

Figure: INTL (2014) traded volume using 5 minute buckets.