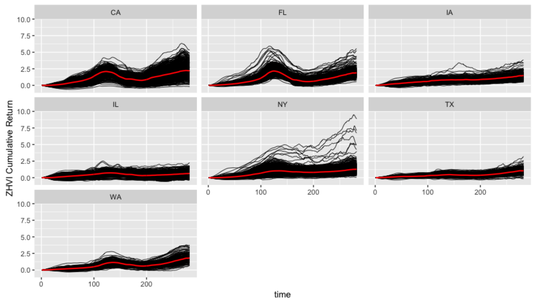
Stat 547 Final Project: Functional Data Analysis for Zillow Home Value Index (ZHVI) Data

Xingche Guo

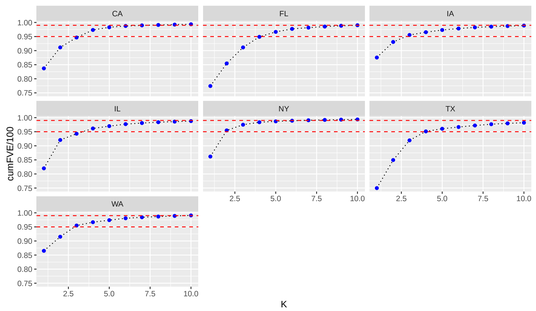
Summary of ZHVI Data in the Analysis

- Monthly measured from 1996-06 to 2019-09 (282 time grids).
- 3685 unique sites located in 7 different states (CA, FL, IA, IL, NY, TX, WA).
- Use transform data: $X^*(t) = \frac{X(t)-X(1)}{X(1)}$ to make two home value in great difference comparable.
- The coordinates (longitude and latitude) of each sites is obtained by R function geocode under API ggmap.

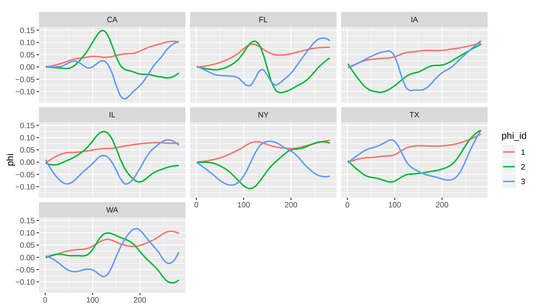


FPCA on ZHVI Data

Cumulated FVE:



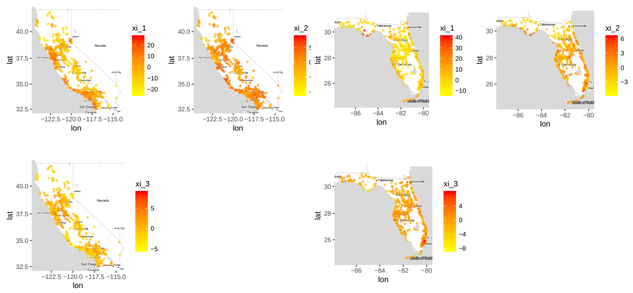
First 3 Eigenfunctions:



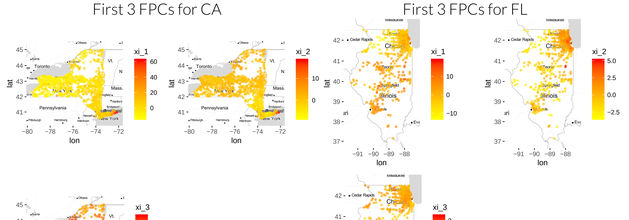
According to the results, FPCA explains the data well, first 3 principle components account for 95% of total variation, the eigenfunctions have similar patterns for each states.

Spatial Effects of FPCs

Project the first 3 FPCs on google map:



First 3 FPCs for CA

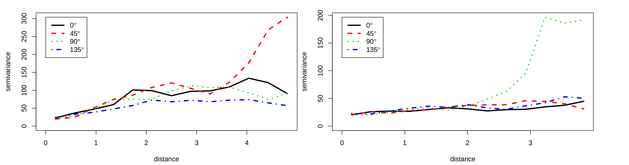


First 3 FPCs for FL

First 3 FPCs for NY

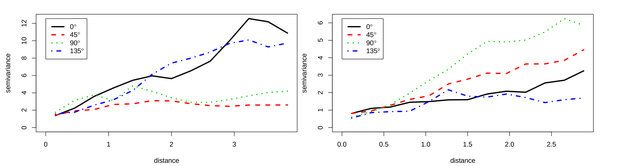
First 3 FPCs for IL

Plot the directional variogram for the 1st FPCs:



Directional Variogram for CA

Directional Variogram for FL



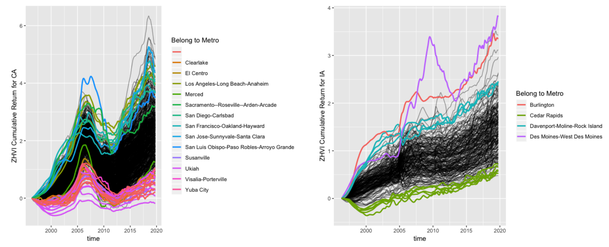
Directional Variogram for NY

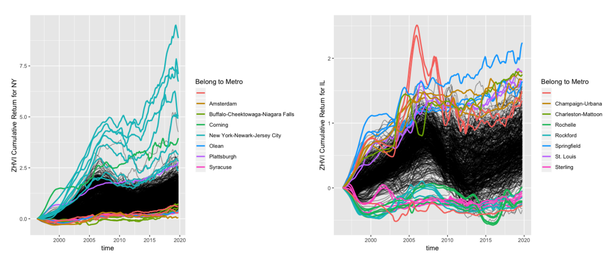
Directional Variogram for IL

- According to the scatterplot and directional variogram plot, the FPCs are anisotropically spatially correlated.
- Should use spatial FPCA on the analysis of spatially correlated functional data.

Check Outliers Using Functional Depth

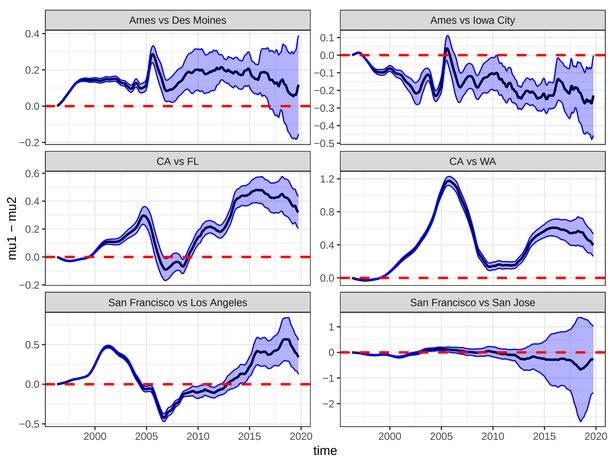
Outlier plots for 4 states:





Inference on ZHVI Data

Simultaneous confidence band for the difference of two mean functions:

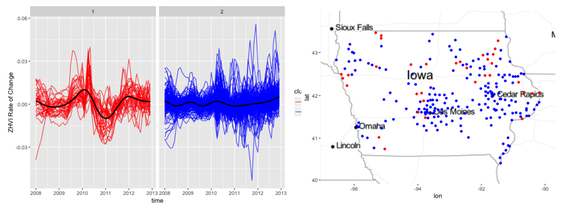


Further Improvements:

- Pre-smooth the curves.
- Account for the spatial correlation within one sample.
- Account for the spatial correlation between two samples.

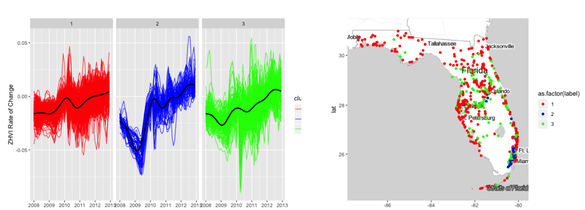
Functional K-means Clustering

- Use transform data: $X^{**}(t) = \frac{X(t)-X(t-1)}{X(t-1)}$ to describe the ZHVI change rate.
- We should repeat the functional K-means algorithm multiple times to reduce the bias cased by the randomness of starting values.



IA

IA

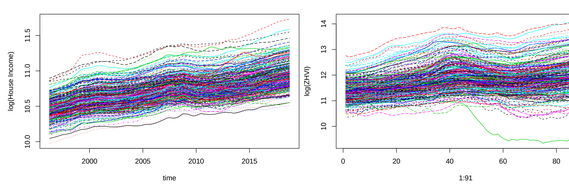


FL

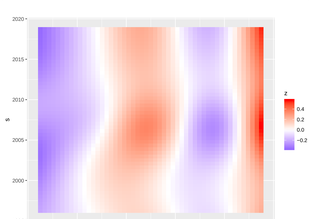
FL

Correlation between Household Income and ZHVI

- Seasonally measured from summer 1996 to winter 2018 (91 time grids) in 356 unique sites.
- Conduct function-on-function regression between X: logarithm of median household income value and Y: logarithm of ZHVI.



$X(s)$ $Y(t)$



$\beta(t, s)$