Aging Curves in Basketball
STOR 890 Project

Alex Wakim

October 16, 2013

Questions of Interest

- Does performance change differently with age for athletic versus unathletic NBA players?
- Does performance change differently with age for Guards versus Forward/Centers?

Measuring Performance

- **Win Shares (WS):** a commonly used metric for performance by a player in a given season. It is a function of the wins of a player's team, offensive ability, and defensive ability.

Determining Athletic vs. Unathletic

- **AthScore:** $AthScore_i = 2V_i' + A_i' + S_i'$
  - V_i' is standardized maximum vertical jump so that it has mean 0 and variance 1.
  - A_i' is standardized inverse agility time so that it has mean 0 and variance 1.
  - S_i' is standardized inverse sprint time so that it has mean 0 and variance 1.
- Players with an AthScore in the top 33% are considered athletic.
- Players with an AthScore in the bottom 33% are considered unathletic.

The Data

- Win Shares for every season for every player who played in the NBA with their first season between 1981-1982 and 2012-2013. (`www.basketball-reference.com`).
  - A player's age during a season is considered to be their age on February 1st of that season.
- AthScore for every player who had their maximum vertical jump, agility time, and sprint time recorded between 2000 and 2012 and is in the DraftExpress database. (`www.draftexpress.com`).

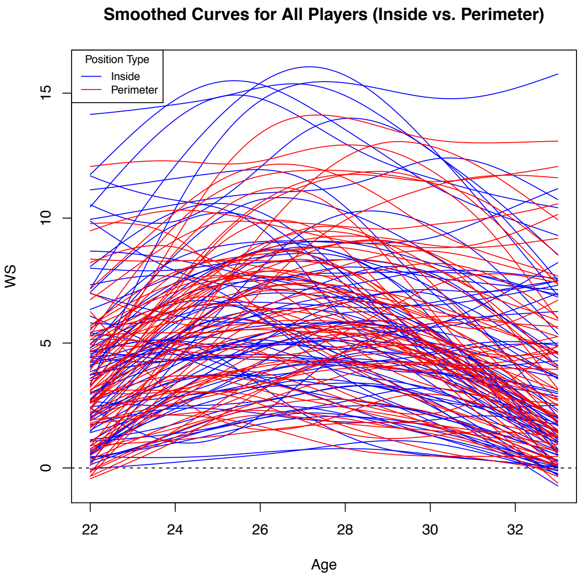
The Data

- **Comparing Position**: must have played every season between ages 22 and 33.
  - 144 players.
- **Comparing Athleticism**: must have played every season between ages 23 and 31.
  - 378 total players; many of which have unknown AthScore.
  - 12 athletic players.
  - 17 unathletic players.

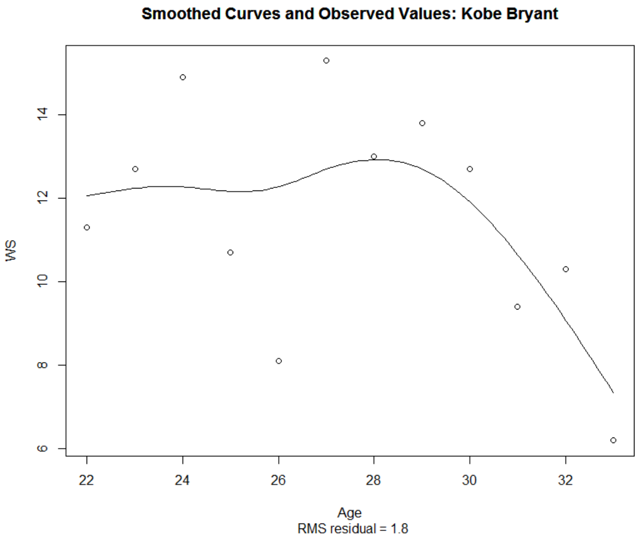
Smoothing Details: Penalized Regression Splines

- Smoothing parameter λ chosen by minimizing GCV.
- **Comparing Position**
  - Cubic b-splines with knots at ages 22, 24, 25, 26, 27, 28, 29, 30, 31, 33.
  - Minimize $\sum_{i=1}^{12} \left(Y_i - \sum_{j=1}^{12} \beta_j B_j(x_i) \right)^2 + 4.46 \int_{22}^{33} \left\{ \sum_{j=1}^{12} \beta_j B_j''(x) \right\}^2 dx$
- **Comparing Athleticism**
  - Cubic b-splines with knots at ages 23, 25, 26, 27, 28, 29, 30, 31.
  - Minimize $\sum_{i=1}^{9} \left(Y_i - \sum_{j=1}^{9} \beta_j B_j(x_i) \right)^2 + 5.74 \int_{23}^{31} \left\{ \sum_{j=1}^{9} \beta_j B_j''(x) \right\}^2 dx$

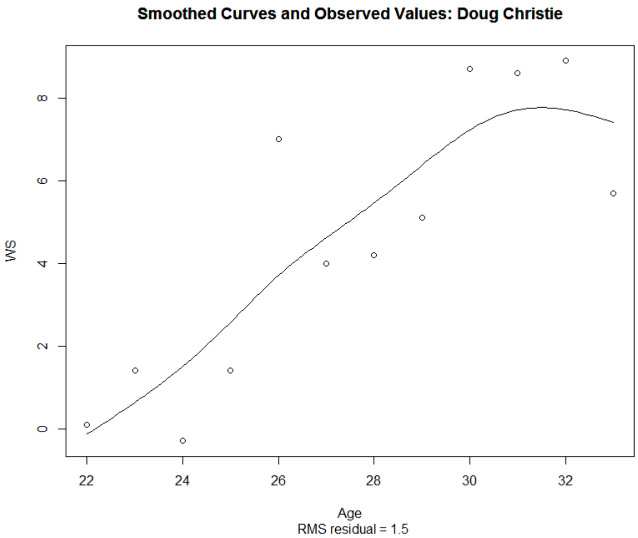
Results: Inside vs. Perimeter



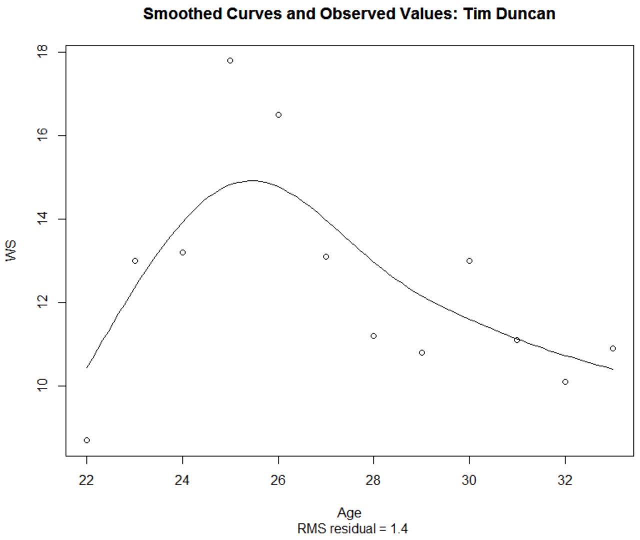
Examples: Smoothed vs. Observed



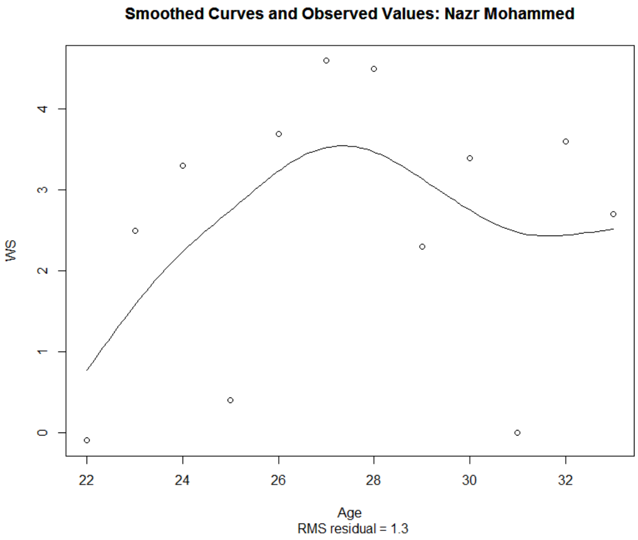
Examples: Smoothed vs. Observed



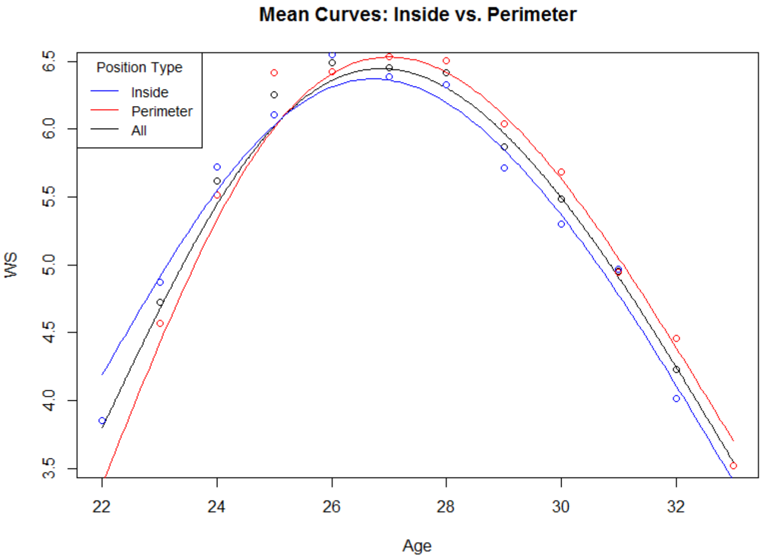
Examples: Smoothed vs. Observed



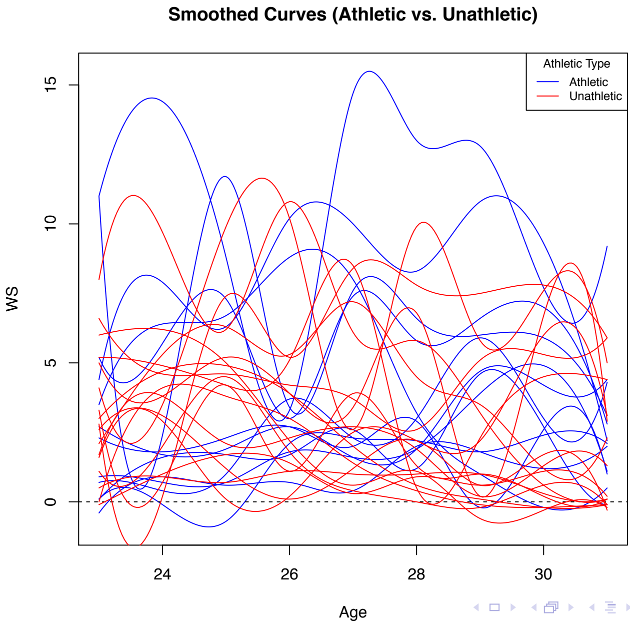
Examples: Smoothed vs. Observed



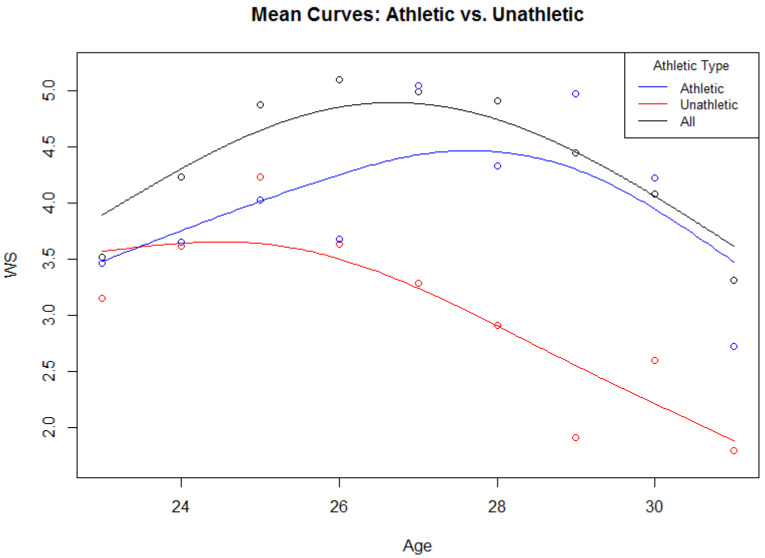
Results: Inside vs. Perimeter



Results: Athletic vs. Unathletic



Results: Athletic vs. Unathletic



Functional Data Analysis of Weight Curves

Gan Liu

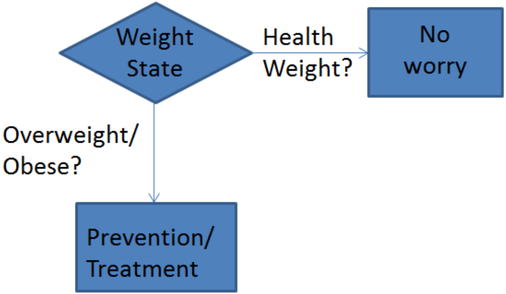
November 4, 2013

Motivation

- More than one-third of U.S. adults (35.7%) are obese;
- Obesity-related conditions include heart disease, stroke, type 2 diabetes and certain types of cancer, some of the leading causes of preventable death;
- The estimated annual medical cost of obesity in the U.S. was $147 billion in 2008 U.S. dollars; the medical costs for people who are obese were $1,429 higher than those of normal weight.
- An adult who has a BMI between 25 and 29.9 is considered overweight.
- An adult who has a BMI of 30 or higher is considered obese.
- BMI=Weight in Kg/(Height in Meters)^2
- For more information, please visit `http://www.cdc.gov/obesity/data/adult.html`

Objectvies

- Get some insights from the weight curves;
- Do some early prediction of weight state of a child;



Data Set

- Height and weight records of 104 children from 2 to 20 years old;
- We divided the data set into two groups, one with 80 records and one with 24 records;
- Data is collected from Prof. Skinner in Public Health Department at UNC;

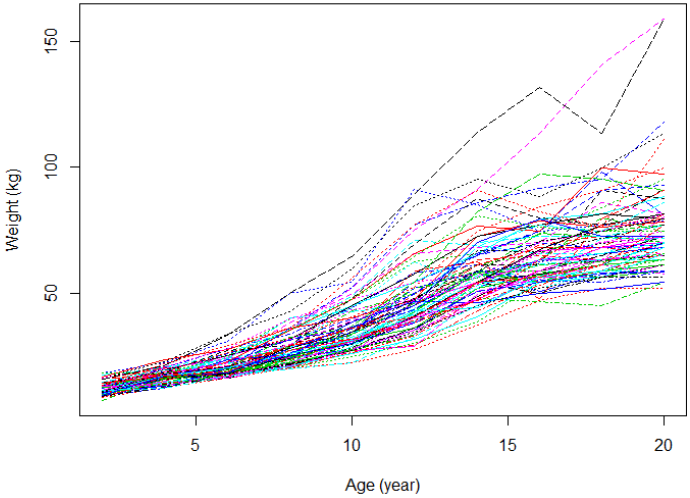


Figure: Original Data

- What if we group the original data based on the BMI of each child at the age of 20?

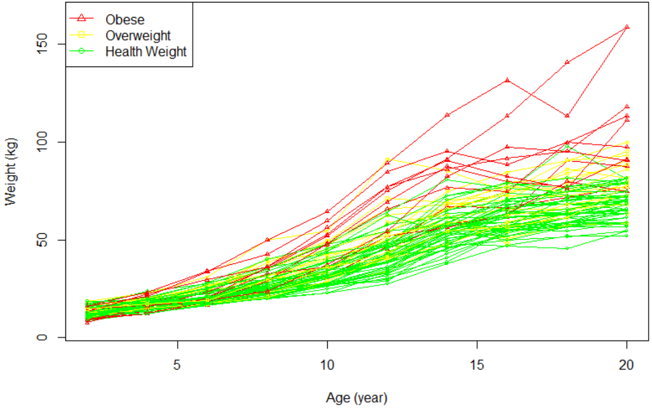


Figure: Classified Data Based on the BMI of Each Child at Age of 20

- For each child, fitting a curve using B-spline basis with the penalty being the integral of the square of the second derivatives;
- Choose the same λ for all of them;
- λ is the one minimizing the average GCV;

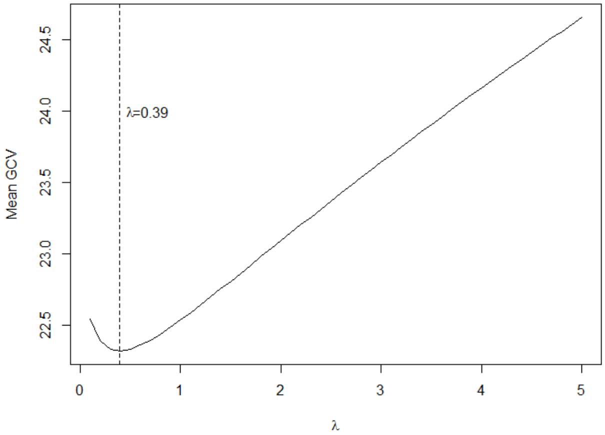


Figure: Mean GCV Plot

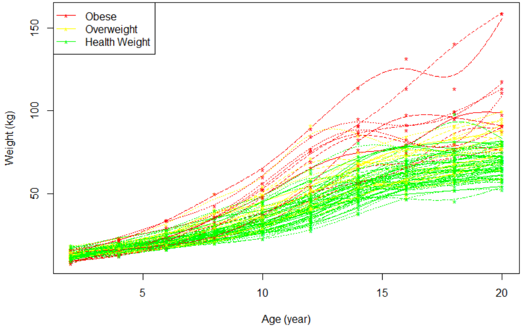


Figure: Curves Obtained by FDA Using B-spline Basis

- Obese children tends to have a overall higher curve;
- Weight curves for overweight children are more or less in the middle;
- Weight curves for health weight children locate in the lower part of the plot;

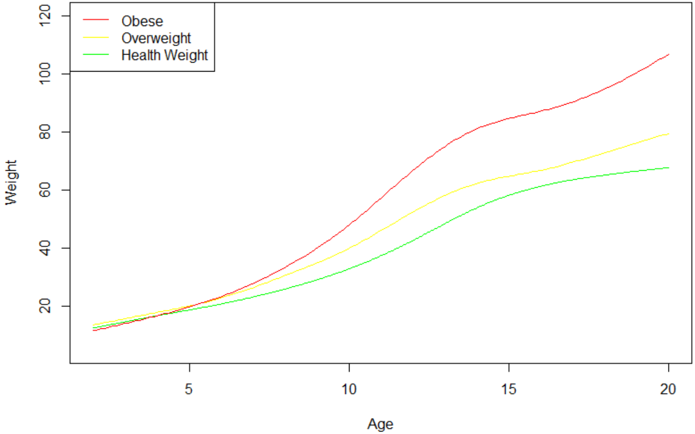


Figure: Mean Curves for Each Group of Children

PCA for Weight Curves

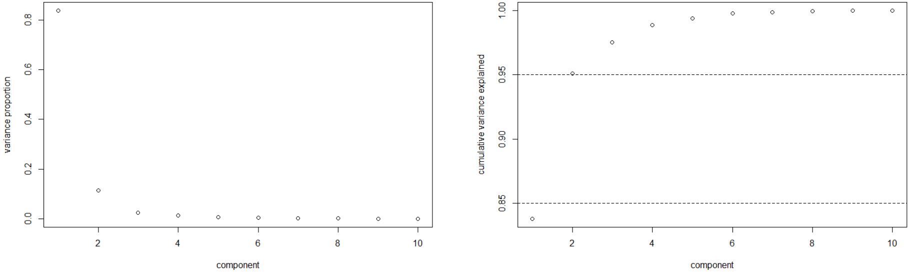


Figure: Variance Proportion of Each Component and Variance Explained

PCA for Weight Curves

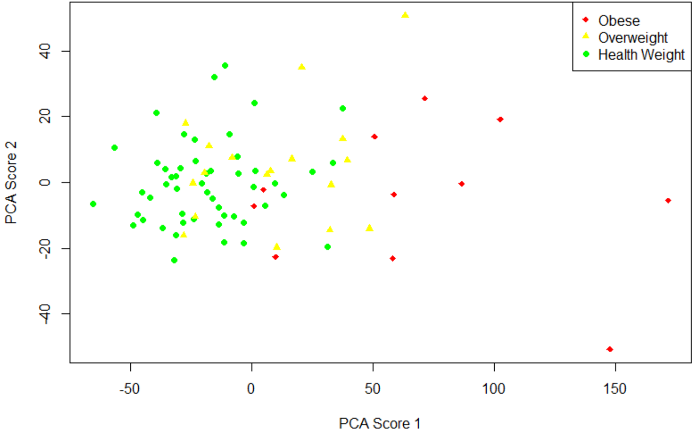


Figure: Plot of PCA Scores

PCA for Weight Curves

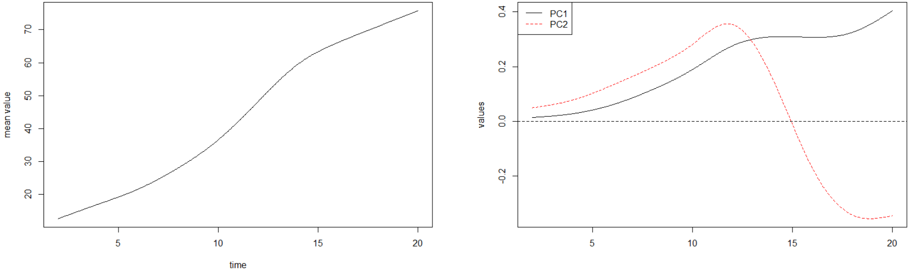


Figure: Mean Curve and First Two PCs

PCA for Weight Curves

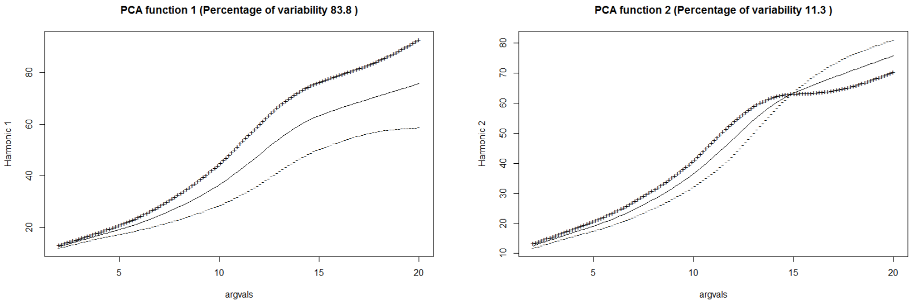


Figure: The two principal component functions or harmonics are shown as perturbations of the mean, which is the solid line. The +'s show what happens when a small amount of a principal component is added to the mean, and the -'s show the effect of subtracting the component.

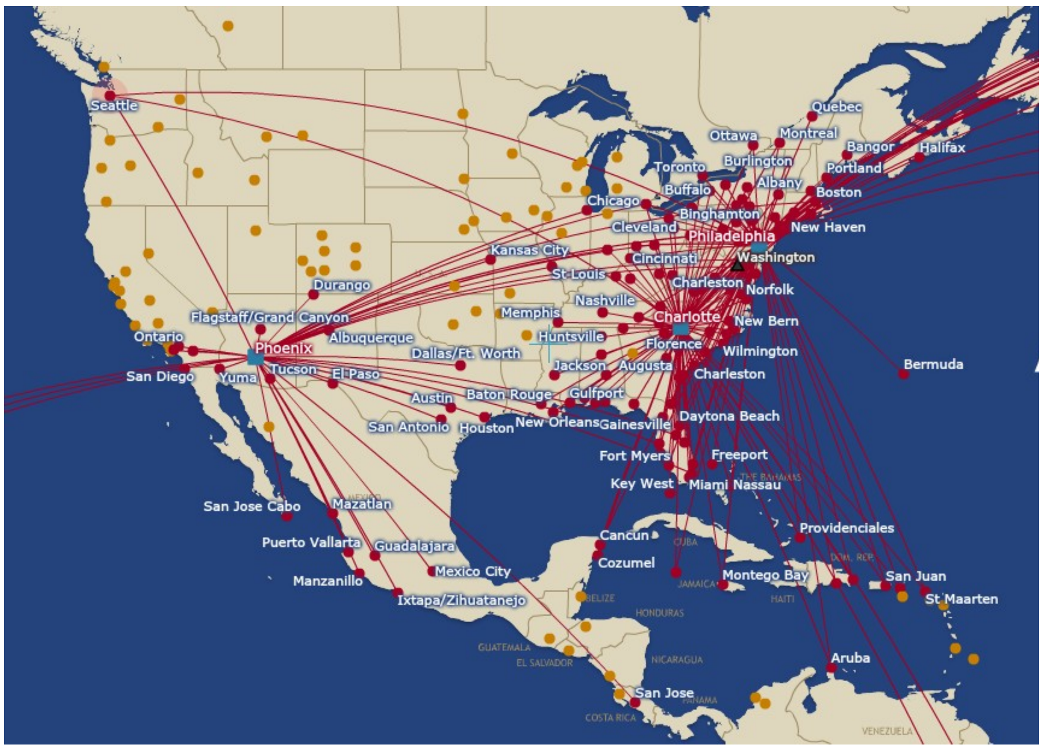
Functional Linear Model: Airline Demand Forecasting

Qing Feng

Nov. 4th ,2013

Background

- Demand
  - Number of passengers willing to fly a specific route at certain time and price



Directional Market

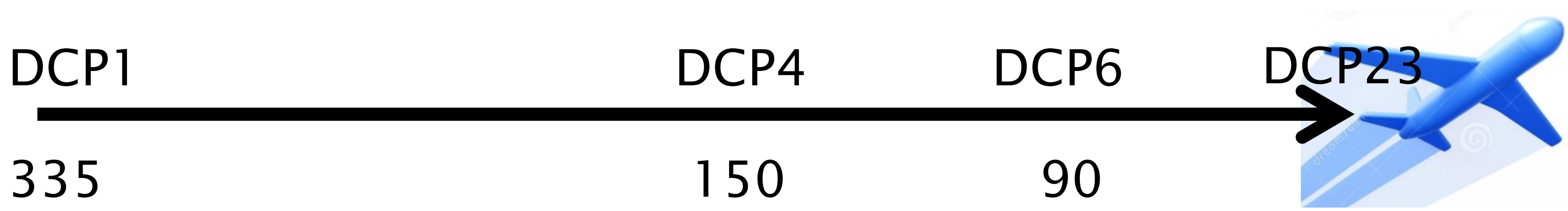
Time Band

Class

- Demand measurement
  - Unconstraint booking count at departure date

Background

- When to make forecast?
  - Booking starts one year before departure
  - Data Collection Point

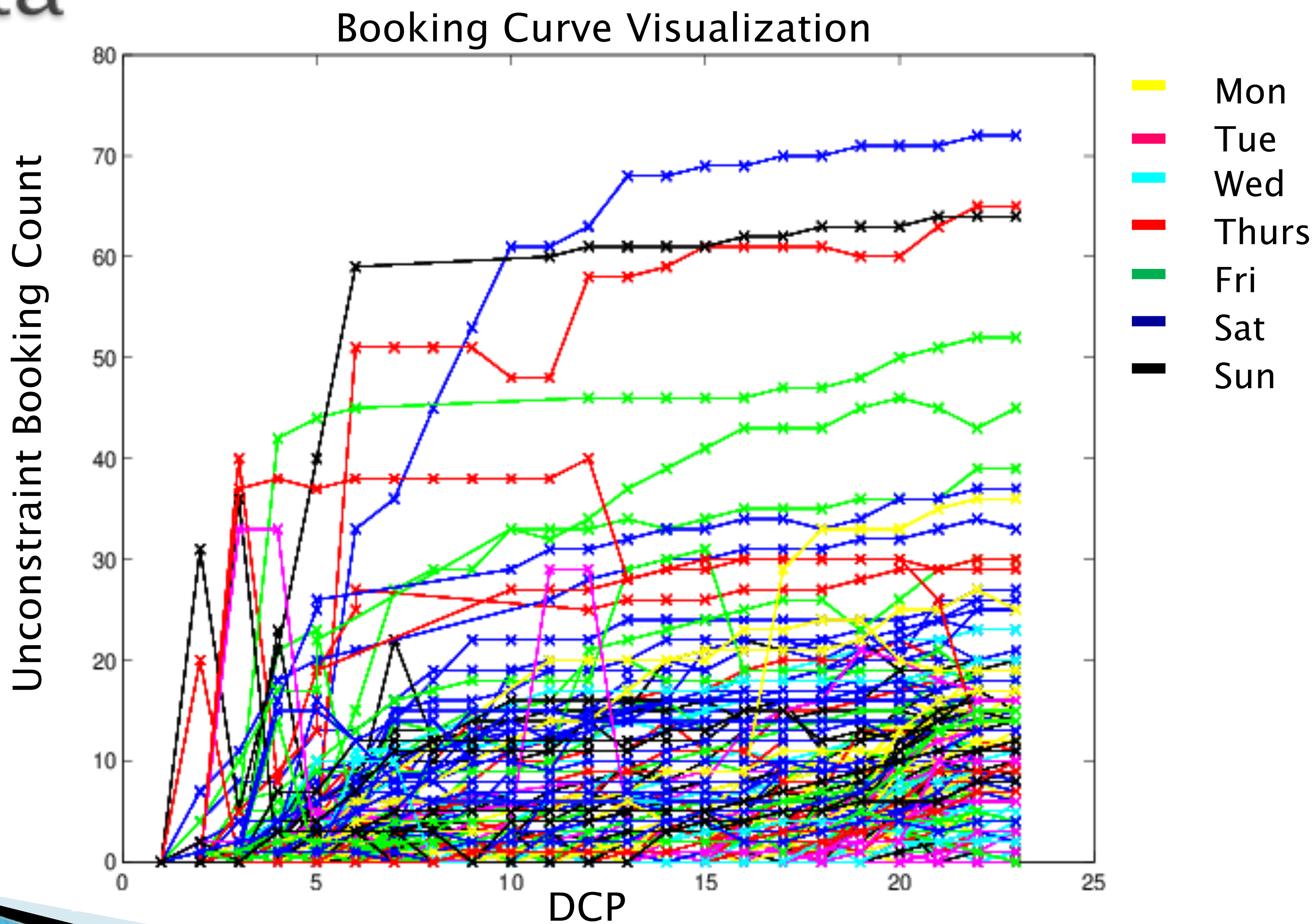


  - Forecast updates at each DCP
- Importance in Revenue management
  - Better fare allocation strategy

Data

- Data scope
  - DTW-CLT directional market
  - Flights departing between 07:35-09:35
  - From 1/3/2012 to 10/17/2012 (207 flights after removal of missing)
  - Class L-Middle level class with relative more booking

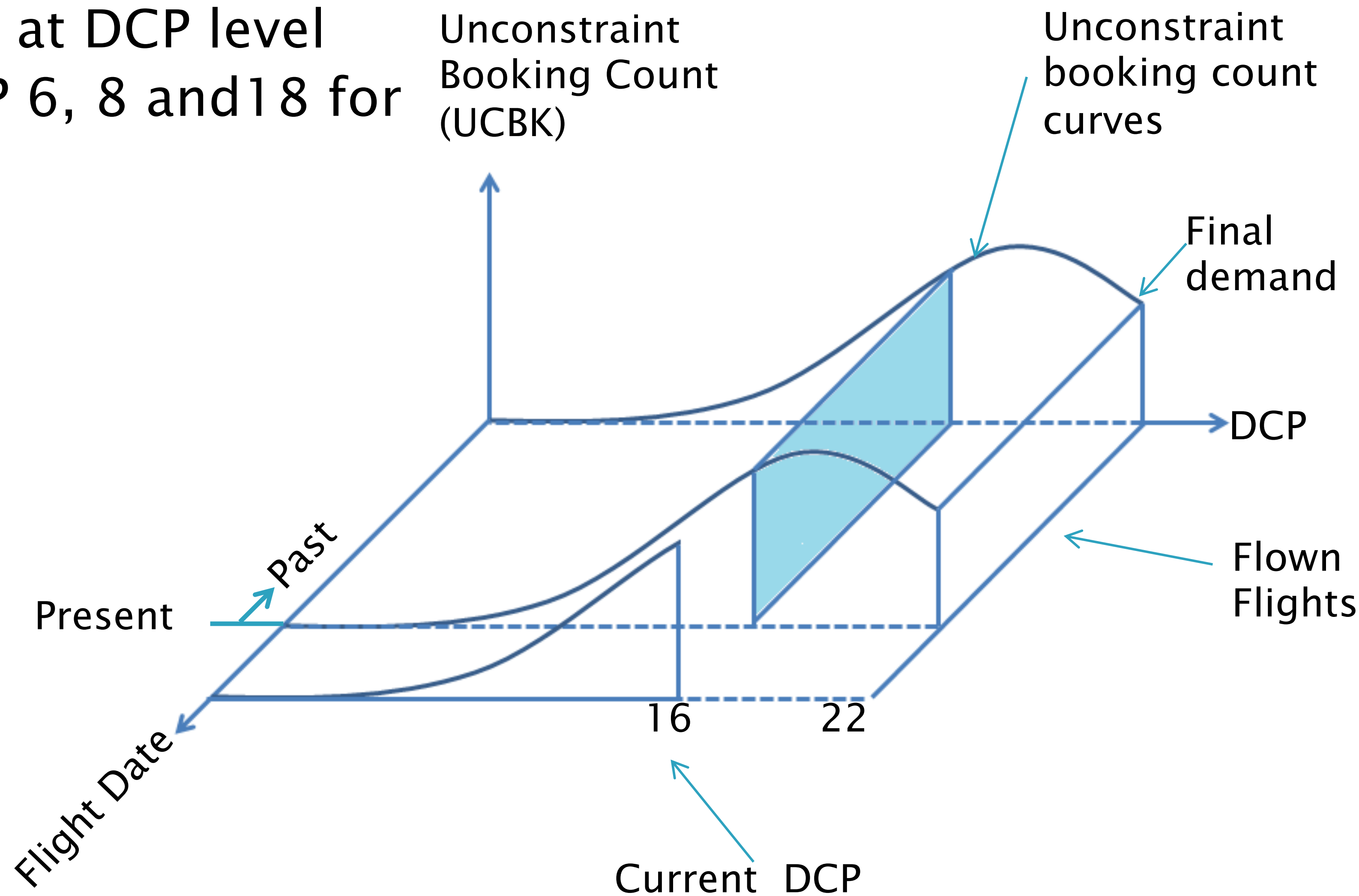
Data



Booking Curve Visualization
Unconstraint Booking Count
DCP
0
10
20
30
40
50
60
70
80
0
5
10
15
20
25
Mon
Tue
Wed
Thurs
Fri
Sat
Sun

Data

- Forecast at DCP level
- Pick DCP 6, 8 and18 for trial



Functional Linear Model

- Data Transformation
  - $\sqrt{Final\ Demand + 0.5}$
  - $\sqrt{Unconstraint\ Booking\ Count + 0.5}$
- Discretized multivariate regression
- Roughness penalized regression
- Comparison

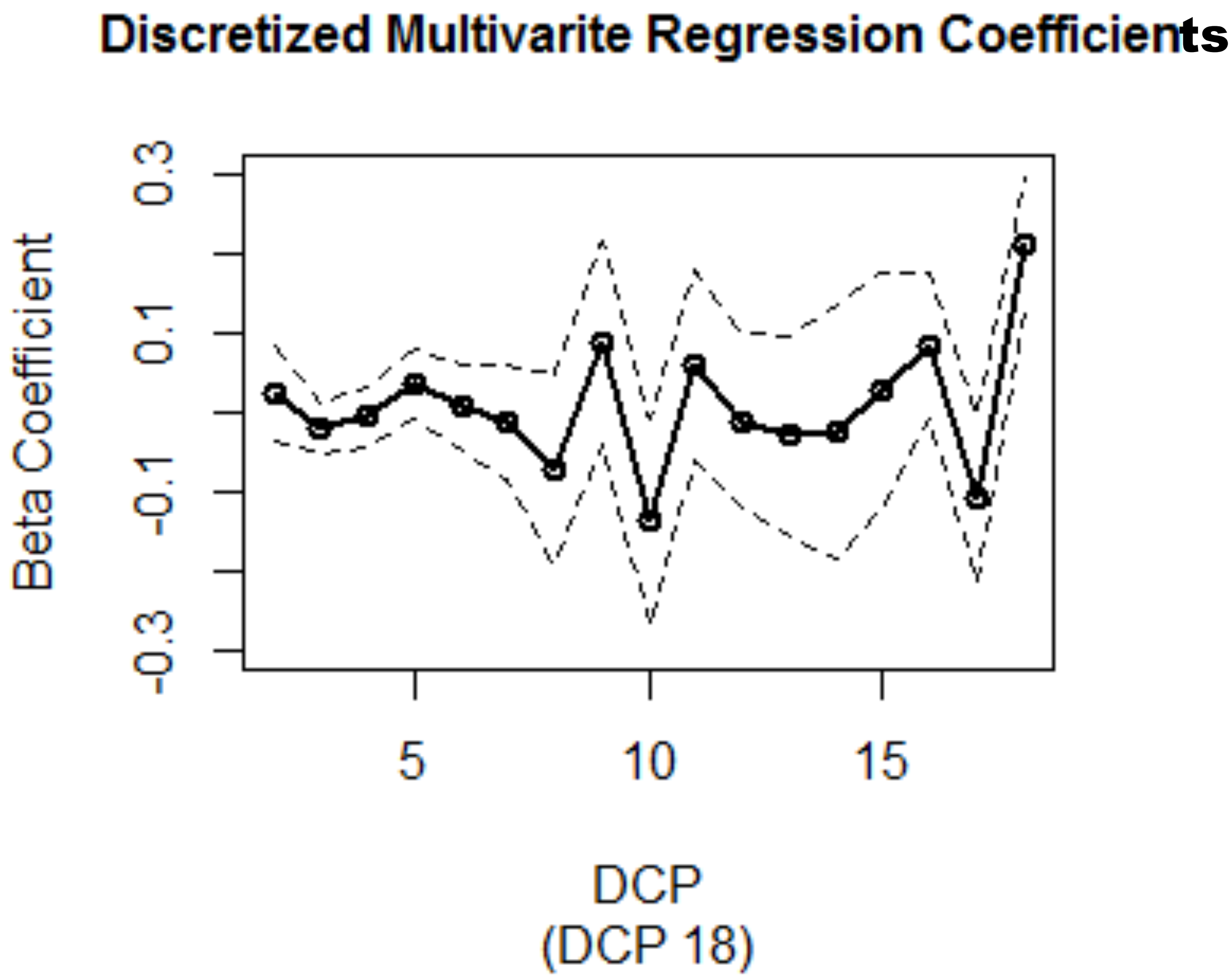
Discretized Multivariate Regression

- Model

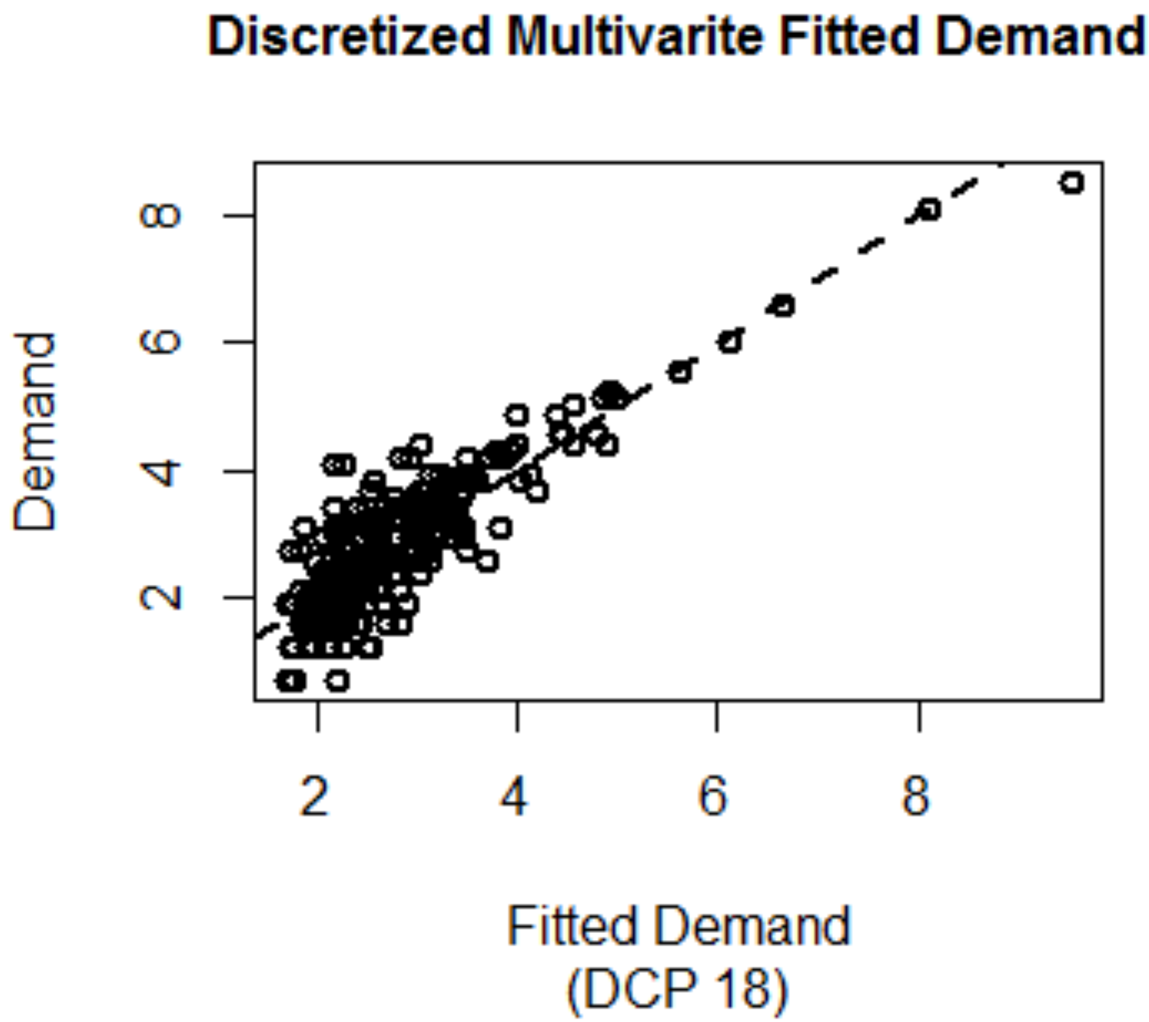
$$Demand_i = \alpha_0 + Dow_i + \sum_{j=1}^{DCP} \beta_j UCBK_j + \varepsilon_i$$

- Forecast is supposed to be made at DCP
- Dow_i is the day of week effect

Discretized Multivariate Regression



Discretized Multivariate Regression



Roughness Penalized Regression

- Model

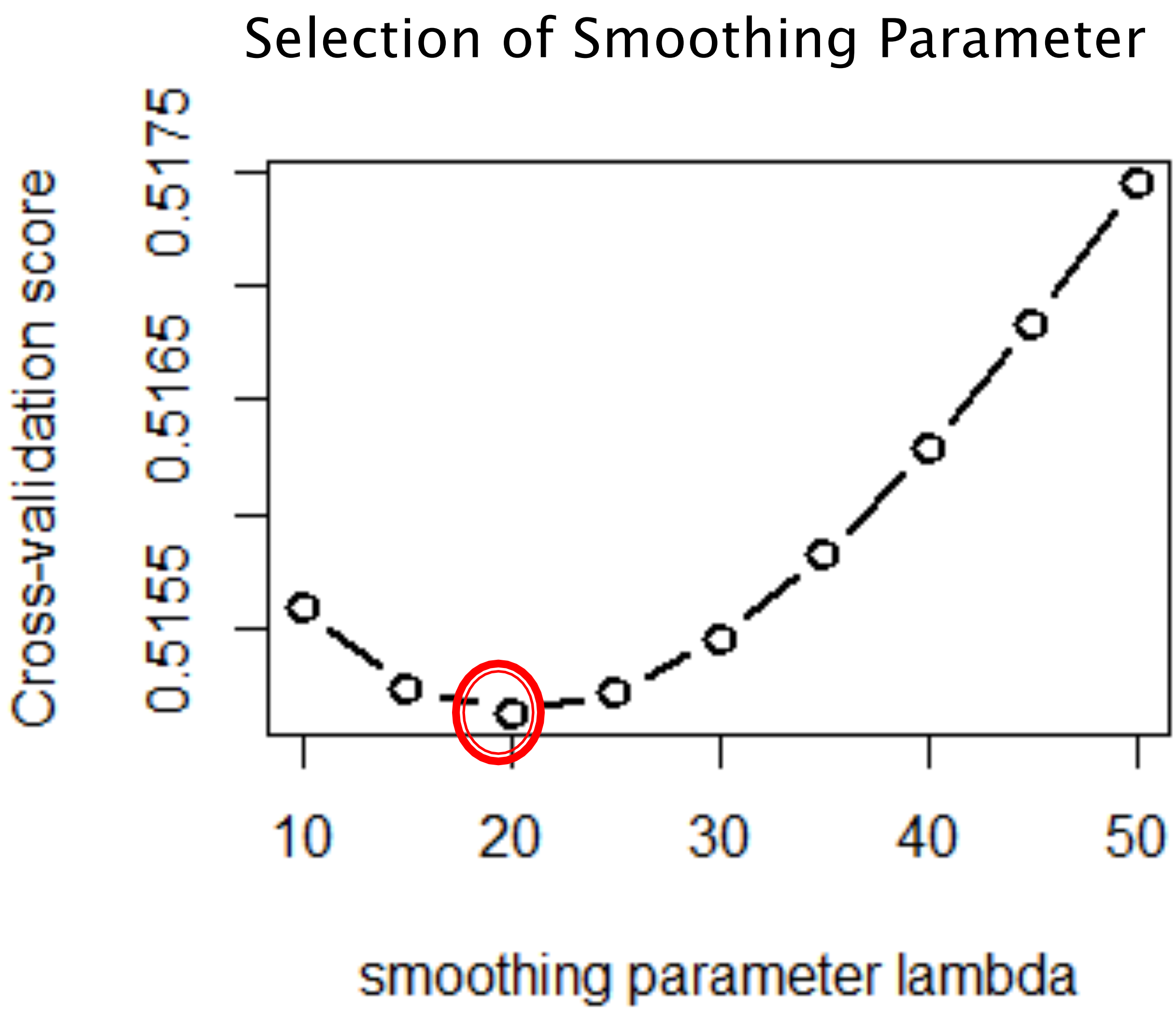
$$Demand_i = \alpha_0 + Dow_i + \int \beta(t) UCBK_i(t) + \varepsilon_i$$

- Dow_i is the day of week effect
- Penalized by

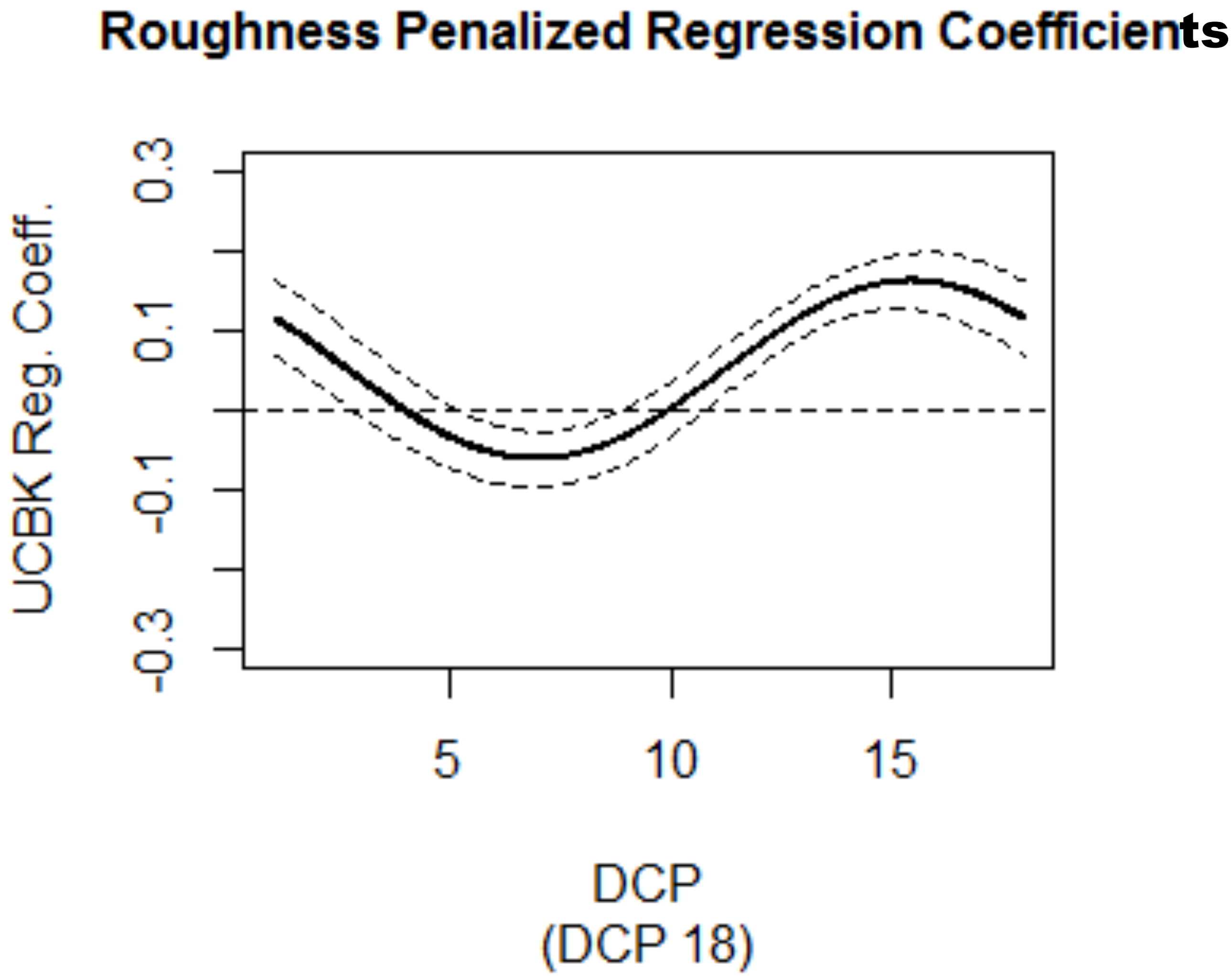
$$\lambda \int [L\beta(t)]^2 dt$$

Where $L\beta = (\omega^2)D\beta + D^3\beta$ is a harmonic acceleration operator

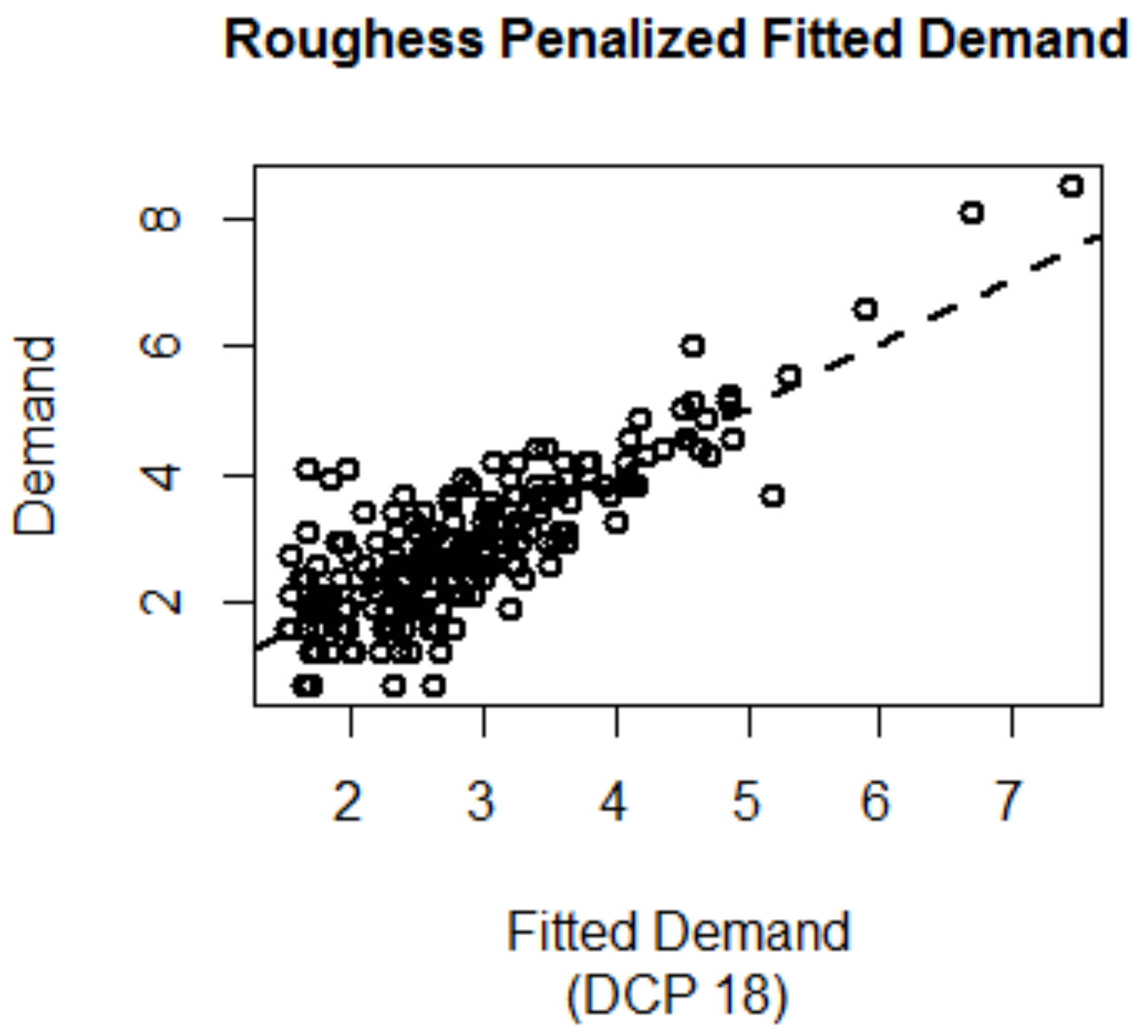
Roughness Penalized Regression



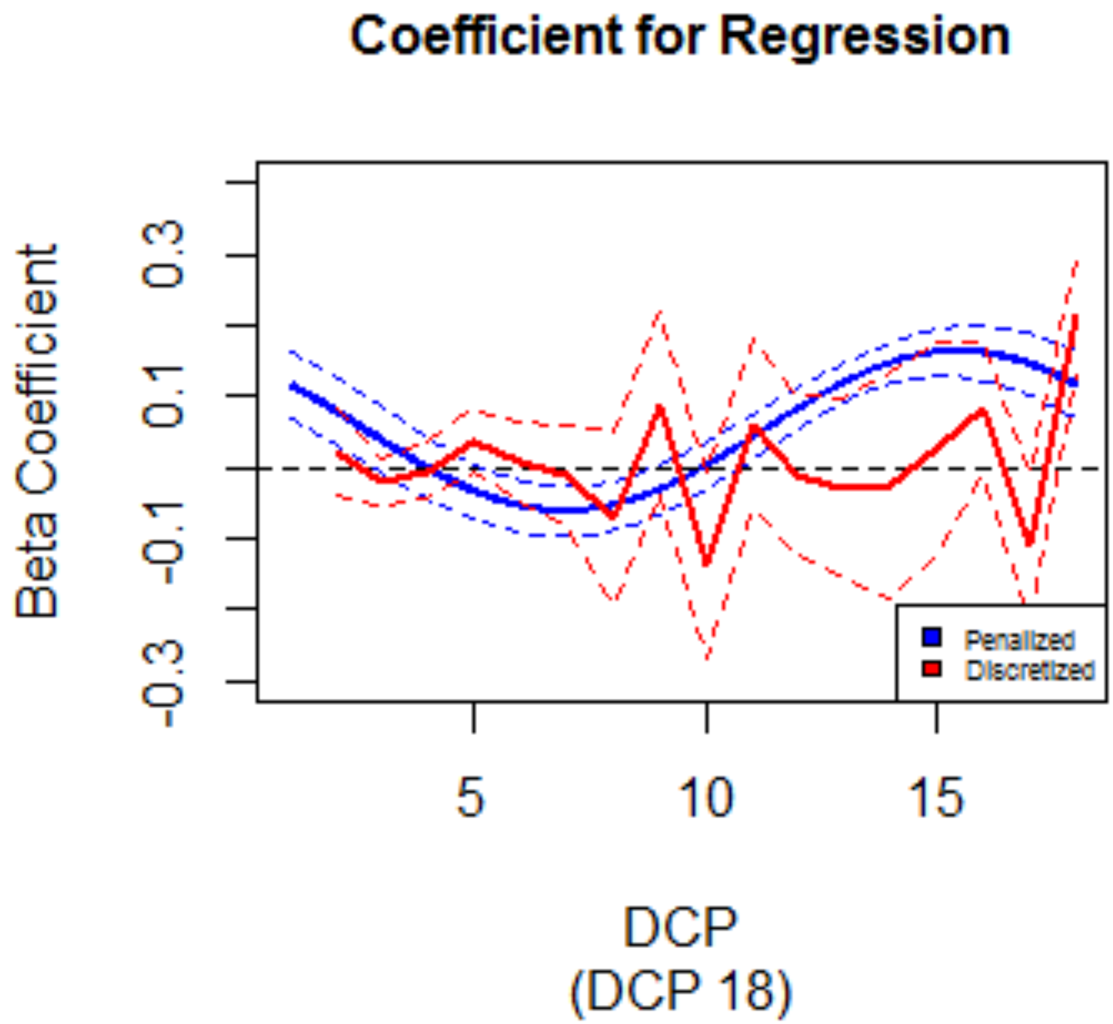
Roughness Penalized Regression



Roughness Penalized Regression



Comparison


Coefficient for Regression
Beta Coefficient
0.3
0.1
-0.1
-0.3
5
10
15
DCP
(DCP 18)
Penalized
Discretized

Summary

- Discretized multivariate tends to over fit
  - Include too much noise at early DCP
  - Safe to use near departure when enough information obtained
- Roughness penalized regression
  - Well capture the booking pattern even when information is little