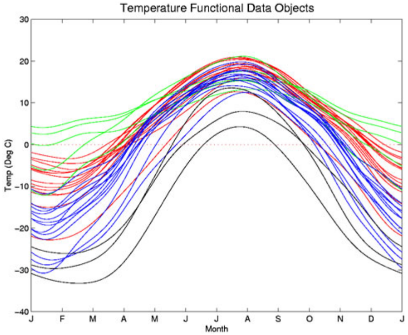
Functional Linear Model

July 2016

Example: Canadian temperature data, smoothed



- Smoothed using 12 Fourier series orthonormal basis over the interval $[0, 12]$
- Red: Atlantic stations
- Blue: Continental
- Green: Pacific
- Black: Arctic

Question can the geographical area explain the annual temperature pattern?

Regression models

$$y_i = \alpha + \mathbf{x}_i \beta + \epsilon_i$$

Response	Covariates	
scalar	vector	
function	vector	Chap. 13
scalar	function	Chap. 15
function	function	Chap. 14 & 16

Regression models with vector covariates

- Analysis of variance: ANOVA
  - response: scalar
- Multivariate analysis of variance: MANOVA
  - response: vector
- Functional analysis of variance: FANOVA
  - response: function
  - predicting temperature curves using climate zones

Functional ANOVA models

- Model: for the mth temperature fn in the gth group,

$$Temp_{mg}(t) = \mu(t) + \alpha_g(t) + \epsilon_{mg}(t)$$

  with constraint

$$\sum_{g=1}^{4} \alpha_g(t) = 0, \quad \text{for all} \quad t$$

- Consider a 35×5 design matrix Z
  - row (mg) for station m in group g
  - this row has 1 in the first column and in column $g + 1$, zero elsewhere
  - denote $Z_{(mg)j}$ as the value in row (mg) and column j

Functional ANOVA models

- Denote function vector $\beta = (\mu, \alpha_1, \alpha_2, \alpha_3, \alpha_4)^T$
- Matrix form:

$$\mathit{Temp} = Z\beta + \epsilon$$

  where *Temp* is the functional vector containing the 35 curves
- The parameter β is now a vector of functions, instead of numbers

Model estimation

- Need to minimize

$$LMSSE(\beta) = \sum_{g=1}^{4} \sum_{m=1}^{N_g} \int [Temp_{mg}(t) - \sum_{j=1}^{5} Z_{(mg)j}\beta_j(t)]^2 dt$$

with constraint

$$\sum_{j=2}^{5} \beta_j = 0$$

- Estimation approaches
  - Pointwise minimization
  - Regularized basis expansion

Example: Canadian temperature data, $\alpha_g(t)$

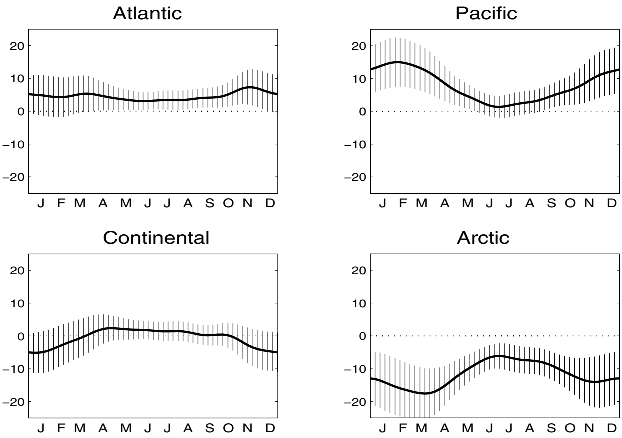


Figure 13.1. The region effects α_g for the temperature functions in the functional analysis of variance model $\texttt{Temp}_{mg}(t) = \mu(t) + \alpha_g(t) + \epsilon_{mg}(t)$. The effects $\alpha_g(t)$ are required to sum to 0 for all t. The cross-hatched areas indicate 95% point-wise confidence intervals for the true effects.

Example: Canadian temperature data, $\mu(t) + \alpha_g(t)$

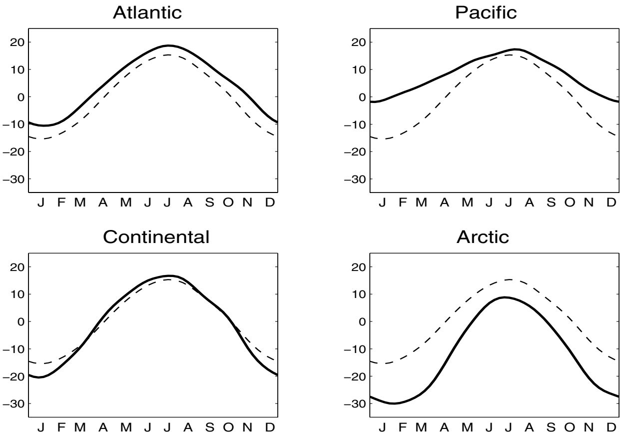


Figure 13.2. The estimated climate zone temperature profiles $\mu + \alpha_g$ for the temperature functions in the functional analysis of variance model (solid curves). The dashed curve is the Canadian mean function μ.

Model diagnostics

- Sum of squared error fn:

$$SSE(t) = \sum_{mg} [Temp_{mg}(t) - Z_{mg}\hat{\beta}(t)]^2$$

- Total sum of squares fn:

$$SSY(t) = \sum_{mg} [Temp_{mg}(t) - \hat{\mu}(t)]^2$$

- Mean squared error fn:

$$MSE(t) = SSE(t)/df(error)$$

  where $df(error)$ is sample size minus num. of independent fn

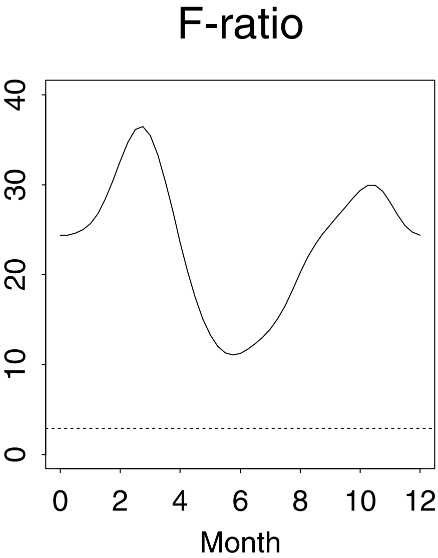
- Mean squares fn for the model:

$$MSR(t) = (SSY(t) - SSE(t))/df(regression)$$

- F-ratio fn:

$$FRATIO(t) = MSR(t)/MSE(t)$$

Example: Canadian temperature data, F-ratio



Computational issues, general model

- Model:
$$y(t) = Z\beta(t) + \epsilon(t)$$
with constraint $L\beta = 0$, Z is $N \times q$

- Fitting criterion:
$$LMSSE(\beta) = \int [y(t) - Z\beta(t)]^T [y(t) - Z\beta(t)] dt$$

- Pointwise minimization:
  - minimize the above criterion for each t individually
  - no restriction on $\beta(t)$ as a fn

Regularized basis expansion approach

- Suppose $y(t) = C\phi(t)$
  - the N-vector y: N observed response fn
  - the K_y-vector ϕ: basis fn (Fourier or B-splines)
  - the $N \times K_y$ matrix C: expansion coefficients
- Basis expansion for the regression fns β:

$$\beta = B\theta$$

  where θ contains the K_β basis fns
- In some cases, we may use $\phi = \theta$

Regularized basis expansion approach

- Fitting criterion:

$$PENSSE(y|\beta) = \int [y(t) - Z\beta(t)]^T [y(t) - Z\beta(t)] dt + \lambda \int (D^2\beta)^T (D^2\beta)$$

- Applying the basis expansion, we have

$$PENSSE(y|\beta) = \int (C\phi - ZB\theta)^T (C\phi - ZB\theta) + \lambda \int (D^2 B\theta)^T (D^2 B\theta)$$

which becomes

$$PENSSE(y|\beta) = trace(C^T C J_{\phi\phi}) + trace(Z^T ZBJ_{\theta\theta} B^T)$$

$$-2trace(BJ_{\theta\phi} C^T Z) + \lambda trace(BRB^T)$$

  - $J_{\phi\phi} = \int \phi\phi^T$, $J_{\theta\theta} = \int \theta\theta^T$, $J_{\phi\theta} = \int \phi\theta^T$
  - $R = \int (D^2\theta)^T (D^2\theta)$

Functional Linear Model with Scalar Response

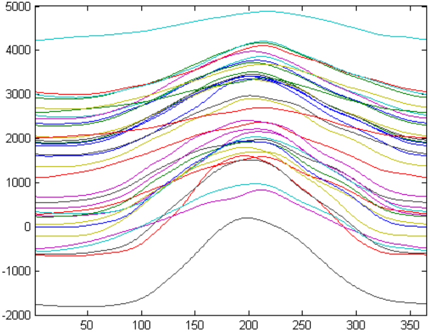
Chap 15 of FDA-RS

Chap 9 of FDA-R-Matlab

July 2016

Models for Scalar Responses

Temperature curves shifted by total annual precipitation:



We want to relate annual precipitation to the *shape* of the temperature profile.

A First Idea

We observe $y_i, x_i(t)$

Choose $t_1, \dots, t_k$

Then we set

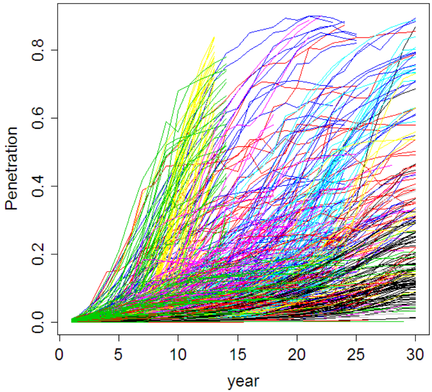
$$\begin{aligned} y_i &= \alpha + \sum \beta_j x_i(t_j) + \epsilon_i \\ &= \alpha + \mathbf{x}_i \beta + \epsilon \end{aligned}$$

And do linear regression.

But how many $t_1, \dots, t_k$ and which ones?

Market Penetration Data

Yearly penetration (%) of consumer products in various countries.



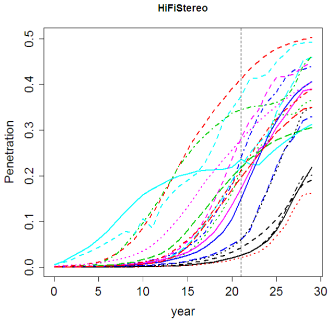
Interest in

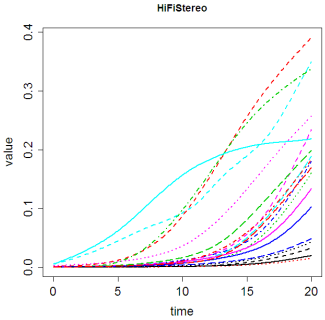
- Forecasting
- Time to “slowdown”
- Differences between countries/products

Sood, A., James, G. and Tellis, G. (2009) "Functional Regression: A New Model for Predicting Market Penetration of

HiFi Stereos

Introduced into 30 Countries in 1977





Attempt to forecast 2007 penetration from 1977-1997 data.

Evaluated at 6 Time Points

```
> btimes = seq(0,20,len=6)
> HiFivals = eval.fd(btimes,HiFifd)
> valmod = lm(y~t(HiFivals))
> summary(valmod)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.20727    0.02389   8.677 9.09e-07 ***
t(HiFivals)1 24.66494   46.88618   0.526 0.607705
t(HiFivals)2 -6.29167   15.28751  -0.412 0.687369
t(HiFivals)3 -2.06591    7.78753  -0.265 0.794951
t(HiFivals)4  5.20007    3.68784   1.410 0.181998
t(HiFivals)5 -4.50342    1.80183  -2.499 0.026620 *
t(HiFivals)6  2.33180    0.52431   4.447 0.000658 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05029 on 13 degrees of freedom
Multiple R-squared: 0.8341,     Adjusted R-squared: 0.7575
F-statistic: 10.89 on 6 and 13 DF,  p-value: 0.0001987
```

In the Limit

If we let $t_1, \ldots$ get increasingly dense

$$y_i = \alpha + \sum \beta_j x_i(t_j) + \epsilon_i = \alpha + \mathbf{x}_i \beta + \epsilon$$

becomes

$$y_i = \alpha + \int \beta(t) x_i(t) dt + \epsilon_i$$

Minimize squared error:

$$\beta(t) = \text{argmin} \sum \left(y_i - \alpha - \int \beta(t) x_i(t) dt \right)^2$$

Identification

Problem:

- In linear regression, we must have fewer covariates than observations.
- if I have $y_i, x_i(t)$, there are *infinitely* many covariates.

$$y_i = \alpha + \int \beta(t)x_i(t)dt + \epsilon_i$$

I can always make the $\epsilon_i = 0$

Hence, the functional view is a necessity, rather than just cosmetic!

Smoothing

We want to insist that $\beta(t)$ is smooth.

Fit by penalized squared error

$$\text{PENSSE}_{\lambda}(\beta) = \sum_{i=1}^{n} \left(y_i - \alpha - \int \beta(t) x_i(t) dt \right)^2 + \lambda \int [D\beta(t)]^2 \, dt$$

Very much like smoothing.

Still need to represent $\beta(t)$ – use a basis expansion

$$\beta(t) = \sum c_i \phi_i(t)$$

Choosing A Basis

Smoothing problem

$$\text{PENSSE}_\lambda(\beta) = \sum_{i=1}^{n} \left(y_i - \alpha - \int \beta(t) x_i(t) dt \right)^2 + \lambda \int [D\beta(t)]^2 \, dt$$

has a unique minimizer over all functions for which $\text{PENSSE}_\lambda(\beta) < \infty$.

- No explicit representation is available.
- However, using a rich enough basis will approximate it well.
- If $x_i(t)$ are represented by a basis, using the same basis often works well.

Calculation

$$
\begin{aligned}
y_i &= \alpha + \int \beta(t)x_i(t)dt + \epsilon_i = \alpha + \left[\int \Phi(t)x_i(t)dt\right]\mathbf{c} + \epsilon_i \\
&= \alpha + \mathbf{x}_i\mathbf{c} + \epsilon_i
\end{aligned}
$$

so

$$
\mathbf{y} = Z \begin{bmatrix} \alpha \\ \mathbf{c} \end{bmatrix} + \epsilon
$$

and

$$
[\hat{\alpha}\ \hat{\mathbf{c}}^T]^T = \left(Z^T Z + \lambda R\right)^{-1} Z^T \mathbf{y}
$$

Then

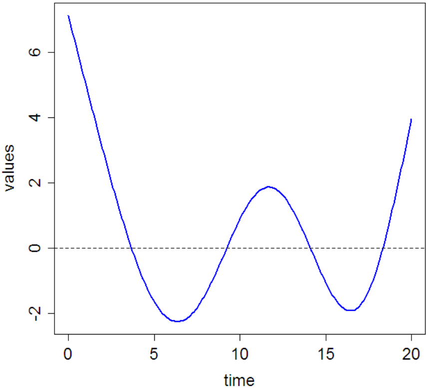
$$
\hat{\mathbf{y}} = \int \hat{\beta}(t)x_i(t)dt = Z \begin{bmatrix} \hat{\alpha} \\ \hat{\mathbf{c}} \end{bmatrix} = S\mathbf{y}
$$

HiFi Penetration

```
> cbasis = create.constant.basis(c(0,20))
> cfd = fd(matrix(1,1,20),cbasis)

> xfdlist = list(cfd,HiFifd)

> betalist = list( fdPar(cbasis,0,0), fdPar(bbasis,2,exp(-8))
> HiFi.reg = fRegress(y,xfdlist,betalist)

> names(HiFi.reg)
[1] "yfdPar"      "xfdlist"     "betalist"    "betaestlist"
[5] "yhatfdobj"   "Cmatinv"     "wt"          "df"

> HiFi.reg$df
[1] 6.175674

> alphahat = HiFi.reg$betaestlist[[1]]$fd$coefs
           [,1]
[1,] 0.01041727
```

HiFi Penetration

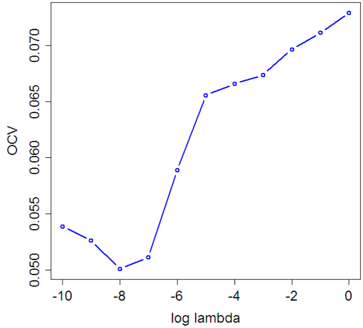
```
plot(HiFi.reg$betaestlist[[2]]$fd)
```



Choosing Smoothing Parameters

Cross-Validation:

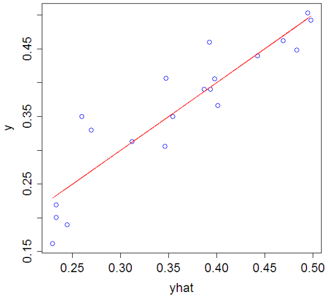
$$\mathrm{OCV}(\lambda) = \sum \left(\frac{y_i - \hat{y}_i}{1 - S_{ii}} \right)^2$$




```
> lambdas = exp(-10:0)
> ocv = rep(0,length(lambdas))

> for(i in 1:16){
>   betalist=list(fdPar(cbasis,0,0),
>       fdPar(bbasis,2,lambdas[i]) )
>   ocv[i]=fRegress.CV(y,xfdlist,betalist)
> }

> plot(-10:0,ocv)
```

Assessing Fit

```
> plot(HiFi.reg$yhatfdobj,y)
> lines(HiFi.reg$yhatfdobj,HiFi.reg$yhatfdobj)
```




```
> sigmae =  sum((y - HiFi.reg$yhatfdobj)^2)/(20-HiFi.reg$df)
> sqrt(sigmae)
[1] 0.0489183
> Rsq = 1 - sigmae/var(y)
[1] 0.7705858
```

Confidence Intervals

Following from smoothing methods we have that

$$\text{Var}\begin{bmatrix} \hat{\alpha} \\ \hat{\mathbf{c}} \end{bmatrix} = \sigma_e^2 \left(Z^T Z + \lambda R\right)^{-1} Z^T Z \left(Z^T Z + \lambda R\right)^{-1}$$

Assuming independent

$$\epsilon_i \sim N(0, \sigma_e^2)$$

Estimate

$$\hat{\sigma}_e^2 = SSE/(n - df), \; df = \text{trace}(S)$$

And confidence intervals are

$$\Phi(t)\hat{\mathbf{c}} \pm 2\sqrt{\Phi(t)^T \text{Var}[\hat{\mathbf{c}}]\Phi(t)}$$

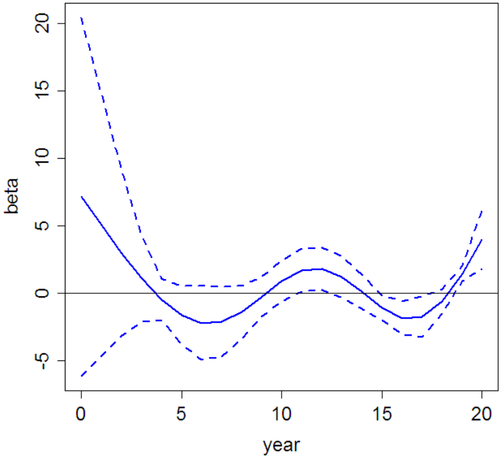
Market Penetration Data

```
> HiFi.reg.std = fRegress.stderr(HiFi.reg,NULL,
      sigmae*diag(rep(1,20)))
> names(HiFi.reg.std)
[1] "betastderrlist" "bvar"           "c2bMap"
```

```
> alphahaterr = HiFi.reg.std$betastderrlist[[1]]$coefs
> c(alphahat-2*alphahaterr,alphahat+2*alphahaterr)

[1] 0.00815757 0.01267696
```

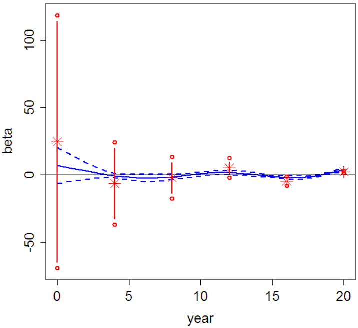
```
> betaest = HiFi.reg$betaestlist[[2]]$fd
> plot(betaest)
> lines(betaest+2*HiFi.reg.std$betastderrlist[[2]])
> lines(betaest-2*HiFi.reg.std$betastderrlist[[2]])
> abline(h=0)
```

Market Penetration Data



In Comparison

```
points(btimes,valmod$coef[2:7])
for(i in 1:6){
ci = c(valmod$coef[i+1]) + c(-2,2)*valmodstd[i+1]
lines(rep(btimes[i],2),ci)
}
```



R^2 comparisons:

0.77 (fRegress)

vs

0.75 (point-wise)

Multivariate and Mixed Functional Linear Regression

What if there are scalar covariates $\mathbf{z}$ and multiple functional covariates $x_1(t), \ldots, x_K(t)$?

$$y_i = \alpha + \mathbf{z}_i\gamma + \sum_{j=1}^{k} \int \beta_j(t) x_{ij}(t) dt + \epsilon_i$$

Then the penalized sum of squares is

$$\sum_{i=1}^{n} \left(y_i - \alpha - \mathbf{z}_i\gamma + \sum_{j=1}^{k} \int \beta_j(t) x_{ij}(t) dt \right)^2 + \sum_{j=1}^{K} \lambda_j \int \left[D_j \beta_j(t) \right]^2 dt$$

Multivariate and Mixed Functional Linear Regression

Calculations

Take

$$\beta_j(t) = \Phi_j(t)\mathbf{c}_j$$

and

$$\zeta = \left[\alpha \; \gamma^T \; \mathbf{c}_1 \; \cdots \; \mathbf{c}_K\right]^T$$

Set

$$\mathbf{x}_{ij} = \int x_{ij}(t)\Phi(t)dt$$

and

$$Z_i = [1 \; \mathbf{z}_i \; \mathbf{x}_{i1} \; \cdots \mathbf{x}_{iK}]$$

Set

$$R_j = \int D_j\Phi_j(t) D_j\Phi_j(t)^T dt$$

$$R = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & \lambda_1 R_1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_k R_K \end{bmatrix}$$

Then

$$\hat{\zeta} = \left(Z^T Z + R\right)^{-1} Z^T \mathbf{y}$$

Multivariate Functional Linear Regression

$$\hat{\zeta} = \left(Z^T Z + R\right)^{-1} Z^T \mathbf{y}$$

Extract $\hat{\alpha}$, $\hat{\gamma}$, $\hat{\mathbf{c}_1}, \ldots, \hat{\mathbf{c}_K}$ from $\hat{\zeta}$.

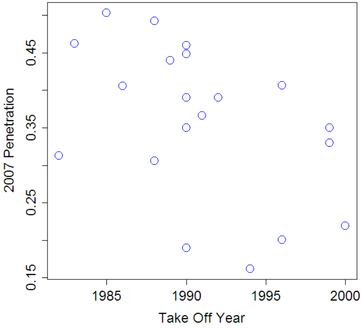
Usual statistics can be calculated:

- Cross-validation estimate
- R^2, error variance
- Co-efficient standard errors

Already used `fRegress` to include intercept in the model.

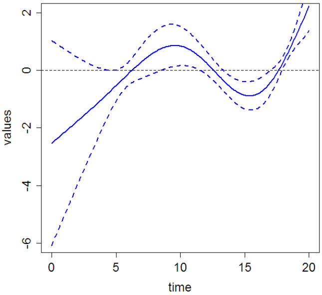
Include “Take-Off” Time

- Measurement of when product penetration sharply increases.
- Functional of covariate function, but *nonlinear*.



Takeoff Time Effect

```
xfdlist[[3]] = fd(yr_takeoff-1977,cbasis)
betalist[[3]] = fdPar(cbasis(0,0))
HiFi.reg2 = fRegress(y,xfdlist,betalist)
```



- Residual std.err: 0.043
- Adjusted R^2: 0.8224657
- $\hat{\alpha} = -0.0005$
- Confidence interval: $[-0.0080,\ 0.0070]$
- $\hat{\gamma} = 0.0005$
- Confidence interval: $[0.0002,\ 0.0009]$

Summary

- Functional linear regression: move from summation to integration
- Identifiability – use a smoothing penalty (remarkable similarity to smoothing)
- Cross validation for smoothing penalty choice
- Usual diagnostic procedures
- Confidence intervals based on residual errors
- Can be generalized to multiple and mixed covariates

Forecasting Time Series of Curves

July 2016

Outline

Background

- Economy - dominated by service sector
- **Call center**: major communication channel

- Other service systems: healthcare delivery systems, ...

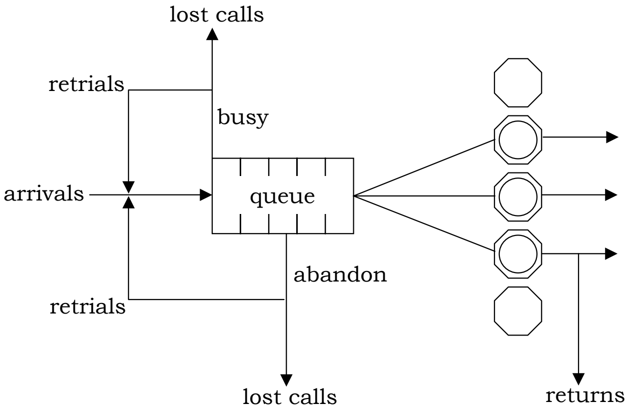
A vast call center world

- 4 million call center agents in US, 800 thousands in UK, 500 thousands in Canada and 500 thousands in India
- Call center costs exceeded $300 billion worldwide
- 70% of the cost for human resource

A wider perspective

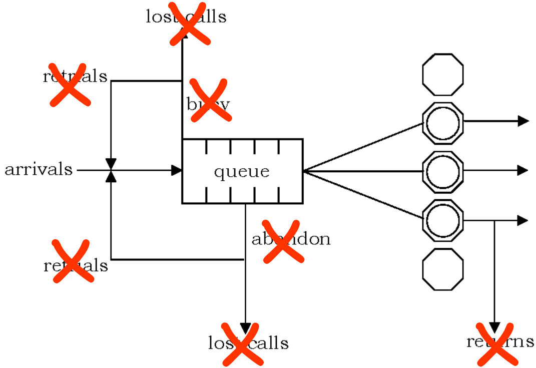
- Multi-disciplinary research area
- Service Enterprise Engineering (NSF)
- Service Science, Management, and Engineering (IBM)
- Service Science - “computer science of the 21st century”

Queueing model for a single call center



Gans, Koole and Mandelbaum (2003)

The $M/M/N + \infty$ model or the Erlang-C model



- no blocking, abandonment, or retrials
- fixed arrival rate λ_j and service rate μ_j for time period j
- exponential inter-arrival and service times

“Standard” model for call center workforce management

1. Forecast offered load (e.g., by the 1/2-hour)

$$\{R_j = \lambda_j/\mu_j : j = 1, \ldots, m\}$$

   where λ_j: arrival rate, μ_j: service rate.

2. Find minimum numbers of agents to make QoS constraint

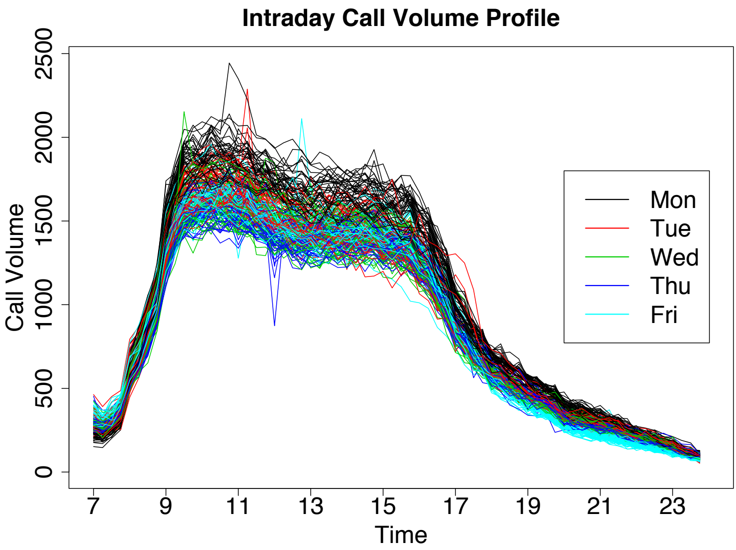
$$s_j = \min\{s \mid \mathrm{P}\{\text{Delay} \leq T\} \geq 1 - \epsilon\}$$

   where in the common 80-20 rule, $T = 20$ seconds and $\epsilon = 20\%$.

3. Find minimum cost assignment of agents to schedules

$$\min\{cN \mid AN \geq s;\ \ N \geq 0;\ \ N \text{ integer}\}$$

   where A: 0-1 schedule matrix, N: # of agents for each schedule, c: schedule cost.


Intraday Call Volume Profile
Call Volume
0
500
1000
1500
2000
2500
7
9
11
13
15
17
19
21
23
Time
Mon
Tue
Wed
Thu
Fri

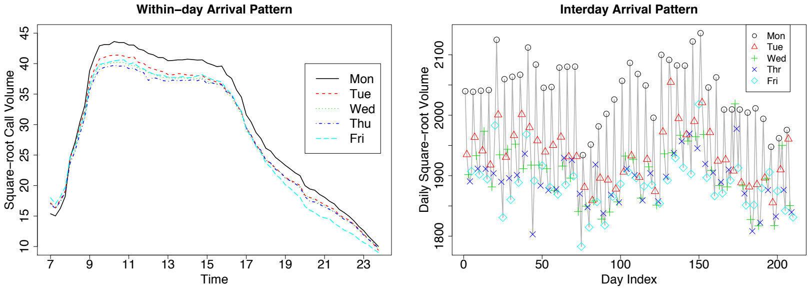
The research problem

- For a given day, call arrivals follow an inhomogeneous Poisson process
  - Day-to-day time series dependence
  - Within-day dependence
  - Seasonal effects
- Two forecasting scenarios:
  - Day-to-day forecasting of future daily arrival rate profile
  - Within-day updating of existing forecast
- Perform stochastic scheduling (with recourse) using the distributional forecasts
- Apply the approach to large-scale real systems

The data

- Bank with a network of 4 call centers in northeast US
- 300K calls/day, 60K/day seeking agents, 1K agents in peak hours
- 210 weekdays between January 6th and October 24th, 2003
- Call volumes during every quarter hour between 7am and midnight
- For each day, 68-dimensional vector of Poisson variables
- Across days, time series of vectors

Two-way arrival pattern



- Smooth intraday trend
- Time series interday dependence

Statistical model

- Consider observations $\{Y_{ij} : j = 1, \ldots, m; \; i = 1, \ldots, n\}$
- Model:

$$Y_{ij} = x_i(t_j) + \epsilon_i(t_j), \qquad j = 1, \ldots, m; \; i = 1, \ldots, n, \qquad (1)$$

  where $x_i(\cdot)$ is an unknown *smooth* curve

- View $\{x_i(t) : 1 \leq i \leq n\}$ as a time series of curves
- Interested in
  - predicting some future smooth curve $x_{n+h}(t)$, $h > 0$
  - updating exist forecast of $x_{n+h}(t)$ based on new information
- FDA usually deals with independent curves

Smooth factor model (SFM)

- Need dimension reduction
- SFM with K factors:

$$x_i(t) = \beta_{i1} f_1(t) + \cdots + \beta_{iK} f_K(t) \tag{2}$$

  - $f_k(t)$: the smooth factor curves
  - $\beta_{i1}, \ldots, \beta_{iK}$: the factor scores
- (1) and (2) together give

$$Y_{ij} = \beta_{i1} f_1(t_j) + \cdots + \beta_{iK} f_K(t_j) + \epsilon_i(t_j) \tag{3}$$

- Consider time series regression model for $\boldsymbol{\beta}_{(i)} = (\beta_{i1}, \ldots, \beta_{iK})^T$
  - $\boldsymbol{\beta}_{(i)}$ are unknown
  - Fit model on the estimated $\boldsymbol{\beta}_{(i)}$
  - Or, estimate the model simultaneously with the factors $f_k(t)$

Model estimation: two-step algorithm

- Estimate $\beta_{(i)}$ and $f_k(t)$ using functional principal component analysis
  - for example, the low-rank-approximation-based approach
  - GCV for penalty parameter selection
  - natural cubic spline interpolation
- Build regression models on $\beta_{(i)}$, to forecast the scores of the future curve
  - time series dependence through lagged responses
  - regression-type dependence on other exogenous covariates

Forecasting future curve

- Denote the first K extracted smooth factors as $\{f_k(t)\}$ and their score series as $\{\boldsymbol{\beta}_k\}$
- Keep the extracted smooth factors as fixed
- Denote the factor score vector and its forecast for the curve $x_{n+h}(t)$ as

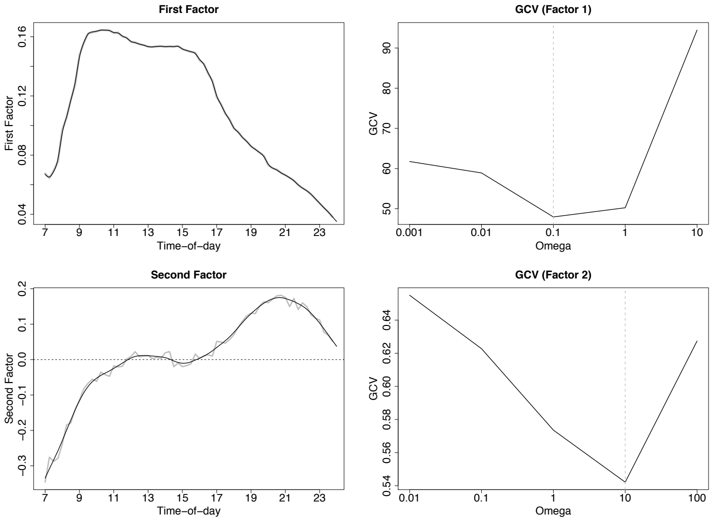
$$\boldsymbol{\beta}_{(n+h)} = (\beta_{n+h,1}, \dots, \beta_{n+h,K})^T$$
$$\hat{\boldsymbol{\beta}}^{\text{SFM}}_{(n+h)} = (\hat{\beta}^{\text{SFM}}_{n+h,1}, \dots, \hat{\beta}^{\text{SFM}}_{n+h,K})^T$$

- Obtain a point forecast of $x_{n+h}(t)$ as

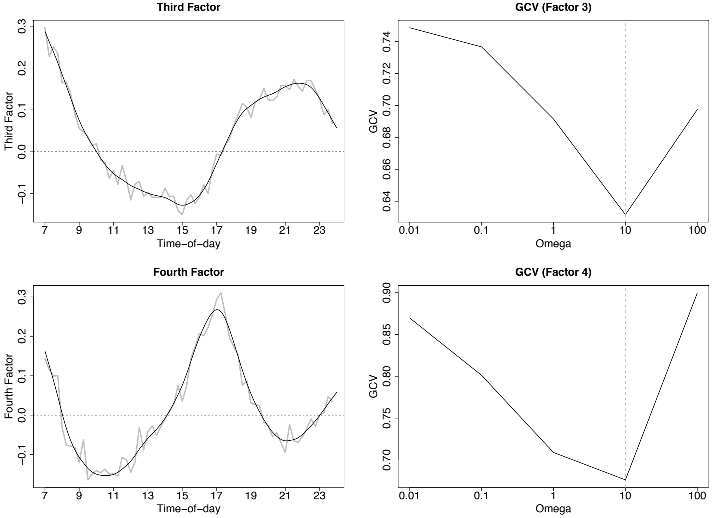
$$\widehat{x}_{n+h}(t) = \hat{\beta}^{\text{SFM}}_{n+h,1} f_1(t) + \cdots + \hat{\beta}^{\text{SFM}}_{n+h,K} f_K(t)$$

- Obtain distributional forecasts for $x_{n+h}(t)$ via bootstrapping the fitted model for $\boldsymbol{\beta}_{(i)}$

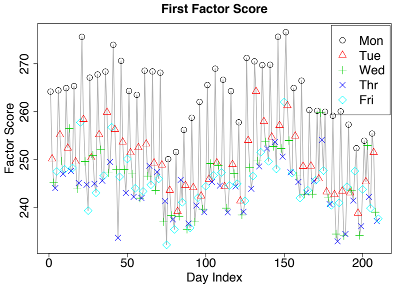
First two intraday factors

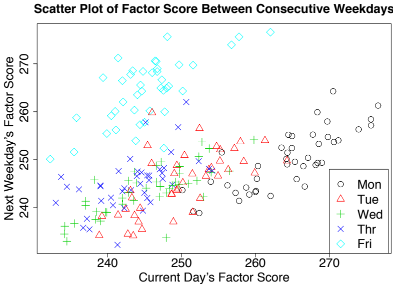


Next two intraday factors



First factor score series β_1





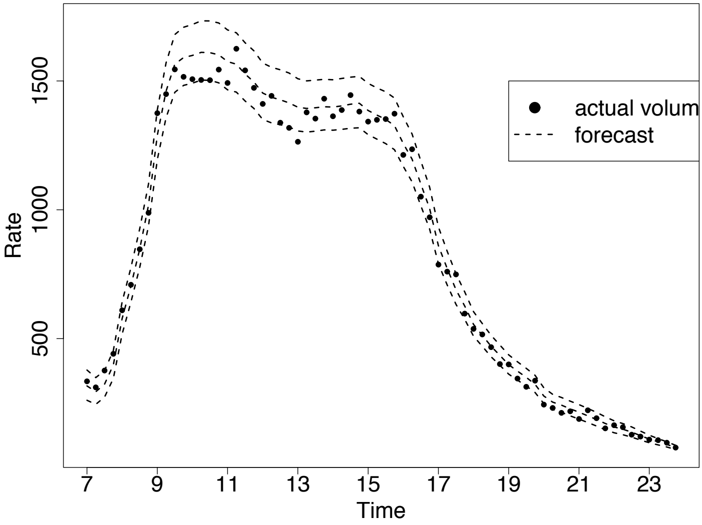
- Varying-coefficient AR(1) model

$$\beta_{i1} = a_1(d_{i-1}) + b_1 \beta_{i-1,1} + \epsilon_{i1},$$

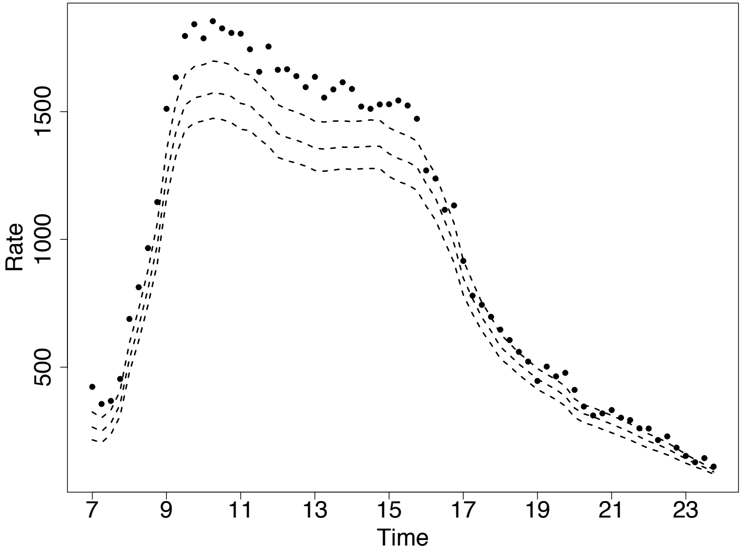
where d_{i-1} denotes the day-of-the-week of day $i-1$

- Similar models for additional factors

Distributional arrival-rate forecasts (often) work well



Night-before forecasts can sometimes be off



Dynamic updating: the problem

- Besides the historical $\{Y_{ij}\}$, also observe

$$\mathbf{y}_{n+1}^{e} \equiv (Y_{n+1,1}, \ldots, Y_{n+1,m_0})^T, \; 1 \leq m_0 < m,$$

  some initial segment of the next curve $x_{n+1}(t)$

- Question: to update the forecast of the *unobserved* latter segment of the curve $x_{n+1}(t)$, denoted as $x_{n+1}^{l}(t)$ for $t > t_{m_0}$

- Two sets of information available
  $\{Y_{ij}\}$ and $\mathbf{y}_{(n+1)}^{e}$

- Regression forecast, $\hat{\beta}_{(n+1)}^{\mathrm{SFM}}$, does not use new info in $\mathbf{y}_{n+1}^{e}$

- To incorporate the new info:
  - treat $\hat{\beta}_{(n+1)}^{\mathrm{SFM}}$ as prior, and shrink $\beta_{(n+1)}$ towards it

Dynamic updating

- Specifically, minimize the following *penalized least squares* (PeLS) criterion with respect to $\boldsymbol{\beta}_{(n+1)}$:

$$\sum_{j=1}^{m_0} \left[Y_{n+1,j} - \{\beta_{n+1,1} f_1(t_j) + \cdots + \beta_{n+1,K} f_K(t_j)\} \right]^2 \qquad (4)$$

$$+\lambda \sum_{k=1}^{K} \left(\beta_{n+1,k} - \hat{\beta}^{\mathrm{SFM}}_{n+1,k} \right)^2 ,$$

where $\lambda > 0$ is a penalty parameter

- In matrix form, the PeLS criterion is

$$\left(\mathbf{y}^e_{n+1} - \mathbf{F}^e \boldsymbol{\beta}_{(n+1)} \right)^T \left(\mathbf{y}^e_{n+1} - \mathbf{F}^e \boldsymbol{\beta}_{(n+1)} \right)$$

$$+\lambda \left(\boldsymbol{\beta}_{(n+1)} - \hat{\boldsymbol{\beta}}^{\mathrm{SFM}}_{(n+1)} \right)^T \left(\boldsymbol{\beta}_{(n+1)} - \hat{\boldsymbol{\beta}}^{\mathrm{SFM}}_{(n+1)} \right) ,$$

where $\mathbf{F}^e$ is a $m_0 \times K$ matrix whose (j,k)-th entry is $f_k(t_j)$, $1 \le j \le m_0$, $1 \le k \le K$

Dynamic updating

- The PeLS update of $\boldsymbol{\beta}_{(n+1)}$:

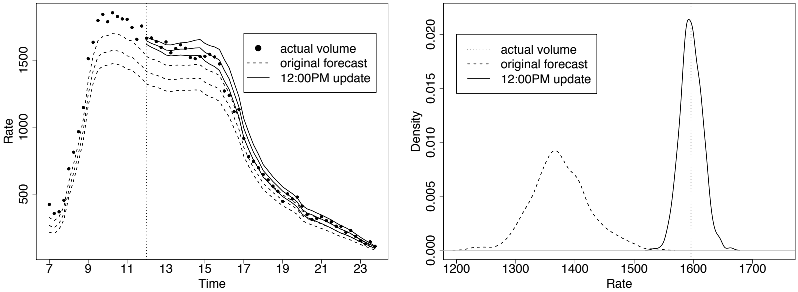
$$\hat{\boldsymbol{\beta}}^{\text{PeLS}}_{(n+1)} = \left(\mathbf{F}^{e^T}\mathbf{F}^{e} + \lambda\mathbf{I}\right)^{-1}\left(\mathbf{F}^{e^T}\mathbf{y}^{e}_{n+1} + \lambda\hat{\boldsymbol{\beta}}^{\text{SFM}}_{(n+1)}\right). \quad (5)$$

- The update of $x^{l}_{n+1}(t)$:

$$\widehat{x}^{\text{PeLS}}_{n+1}(t) = \hat{\beta}^{\text{PeLS}}_{n+1,1} f_1(t) + \cdots + \hat{\beta}^{\text{PeLS}}_{n+1,K} f_K(t), \qquad t > t_{m_0}. \quad (6)$$

- λ selected through minimizing out-of-sample forecasting error
- Update the bootstrapped forecast distribution

Forecast updates can significantly reduce error and uncertainty



Rolling out-of-sample forecast comparison: benefit of updating

	RMSE		
	SFM4	PeLS4 10:00AM	PeLS4 12:00PM
Q1	35.53	33.76	31.81
Median	42.78	40.32	35.80
Mean	51.81	45.95	43.03
Q3	59.64	47.80	41.91

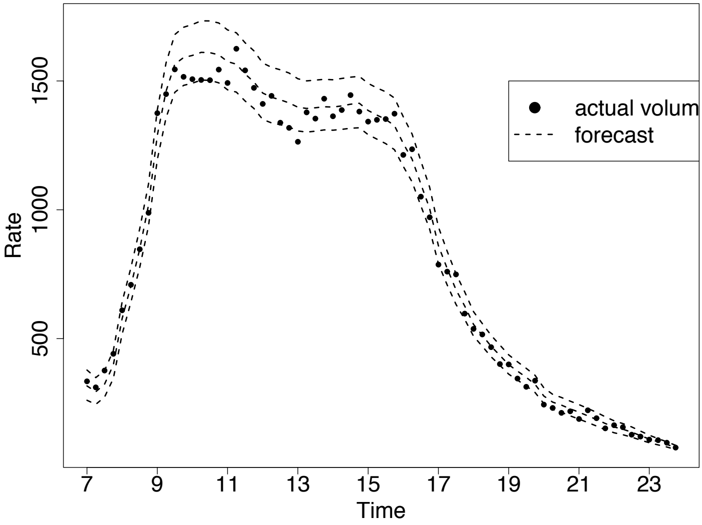
Rolling out-of-sample forecast comparison: benefit of updating

	Average Interval Width		
	SFM4	PeLS4 10:00AM	PeLS4 12:00PM
Q1	121.4	94.0	76.7
Median	125.9	102.8	81.6
Mean	125.9	102.8	81.7
Q3	130.7	108.1	85.7

Our goal

- Develop distributional forecasts for arrival rates
  - update given additional information
- Perform stochastic scheduling using the distributional forecasts
  - recourse actions after forecast updating
  - adjust staffing assignments
    - send agents home early ... $\rightarrow$ reduce cost
    - call in part-time agents ... $\rightarrow$ better achieve QoS measure
- Test the approach in large-scale real systems

Distributional arrival-rate forecasts



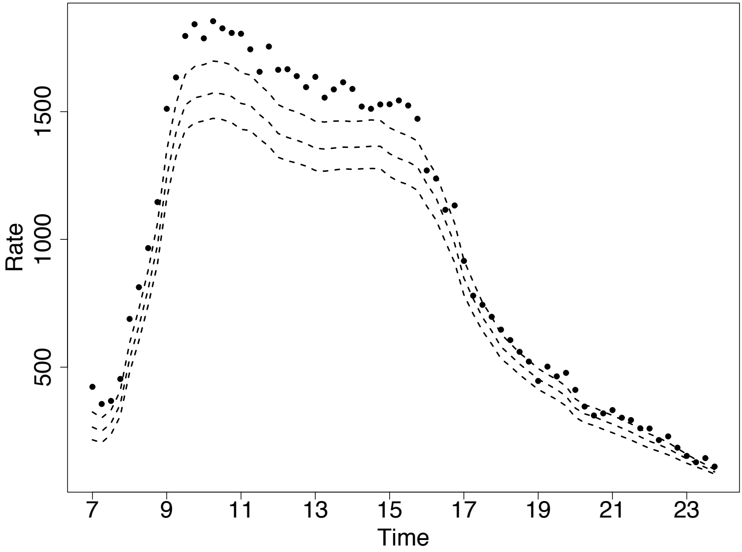
Stochastic program for scheduling agents

- Distribution of the Λ_t's determined from the forecast, $t = 1, \ldots, T$
- α^*: upper bound on expected abandonment rate
- y: # of agents on each of the possible schedules, with cost c
- $a_t y$: # of agents working in period t
- With distributions for Λ_t's, solve the stochastic program

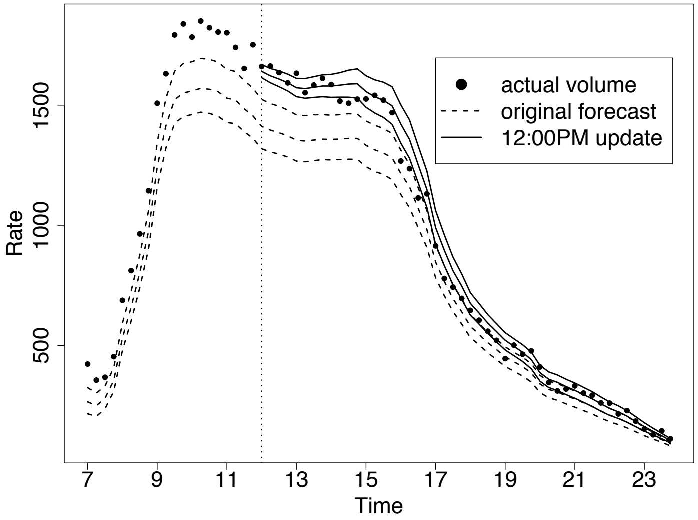
$$\begin{aligned}
&\min \{cy\} \\
&\text{s.t.} \\
&\sum_{t=1}^{T} \mathrm{E}_{\Lambda_t}\left[f(\Lambda_t, a_t y)\right] \le \alpha^* \sum_{t=1}^{T} \mathrm{E}\left[\Lambda_t\right] \\
&y \ge 0; \quad y \text{ integer},
\end{aligned}$$

where $f(\Lambda_t, a_t y)$ is abandonment count in period t

Night-before forecasts can sometimes be off



Forecast updates significantly reduce error and uncertainty



Stochastic programming with recourse - Stage 1

- Solve the same stochastic program

$$
\begin{aligned}
&\min \{cy\} \\
&\text{s.t.} \\
&\sum_{t=1}^{T} \mathsf{E}_{\Lambda_t}\left[f(\Lambda_t, a_t y)\right] \le \alpha^* \sum_{t=1}^{T} \mathsf{E}\left[\Lambda_t\right] \qquad\qquad t = 1, \dots, T \\
&y \ge 0; \quad y \text{ integer}
\end{aligned}
$$

- Calculate

$$
\alpha_I = \frac{\sum_{t=T_0+1}^{T} \mathsf{E}_{\Lambda_t}\left[f(\Lambda_t, a_t y)\right]}{\sum_{t=T_0+1}^{T} \mathsf{E}\left[\Lambda_t\right]},
$$

expected abandonment rate over late part of the planning horizon

Stochastic programming with recourse - Stage 2

- Updated forecast Λ_t^*
- Adjust staffing assignments
  - update original staffing, y, to updated vector z
  - cost of change, $d(y - z)$
- Solve stochastic program with recourse

$$\min \{cy + d(y - z)\}$$

s.t.

$$\sum_{t=T_0+1}^{T} \mathrm{E}_{\Lambda_t^*}\left[f(\Lambda_t^*, a_t z)\right] \leq \alpha_l \sum_{t=T_0+1}^{T} \mathrm{E}\left[\Lambda_t^*\right]$$

$$z \geq 0; \; z \text{ integer}$$

Recourse program that uses 2-stage forecast

- Idea: account for recourse actions in initial schedule
- Example:
  - it costs little to send agents home after T_0
  - it costs a lot to increase staffing after T_0
  - then, should initially staff high and send people home, if necessary
- In two-stage program:
  - 1st-stage periods as before: initial staffing y fixed across scenarios
  - 2nd-stage periods more complex: for each initial scenario, 2nd-stage action z varies

Recourse program that uses 2-stage forecast

- Initial staffing vector y, and random update vector z
- Stochastic program with recourse

$$
\begin{aligned}
&\min \left\{ cy + \mathrm{E}_{\Lambda_1,\ldots,\Lambda_{T_0}} \left[d(y-z) \right] \right\} \\
&\text{s.t.} \\
&\sum_{t=1}^{T_0} \mathrm{E}_{\Lambda_t} \left[f(\Lambda_t, a_t y) \right] \\
&\quad + \sum_{t=T_0+1}^{T} \mathrm{E}_{\Lambda_1,\ldots,\Lambda_{T_0}} \left[\mathrm{E}_{\Lambda_t|\Lambda_1,\ldots,\Lambda_{T_0}} \left[f(\Lambda_t, a_t z) \right] \right] \leq \alpha^* \sum_{t=1}^{T} \mathrm{E}\left[\Lambda_t \right] \\
&y, z \geq 0;\ \ y, z \text{ integer}
\end{aligned}
$$

We test six scheduling schemes

- Two schemes with no updating
  - one scenario = IP ♢
  - 100 scenarios = SP100 ♦
- Two schemes with an afternoon update of the original schedule
  - one scenario = UP □
  - 100 scenarios = UP100 ■
- Two schemes that update an original schedule with recourse
  - one scenario = RP ○
  - 100 scenarios = RP100 ●

Testing the value of the scheduling schemes

1. Preliminary forecast using previous n days of data
2. Solve 4 scheduling problems based on initial forecast
   - IP ♢ and SP100 ♦
   - 1st phase of RP ○ and RP100 ●
3. Update forecast based on 1st part of day
4. Update solutions based on revised forecast
   - IP $\Rightarrow$ UP □ and SP100 $\Rightarrow$ UP100 ■
   - 2nd phase of RP ○ and RP100 ●
5. Simulate using schedules and actual arrival counts

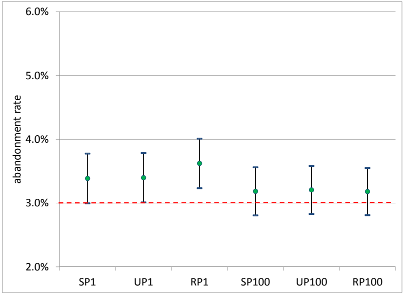
1st set of empirical tests

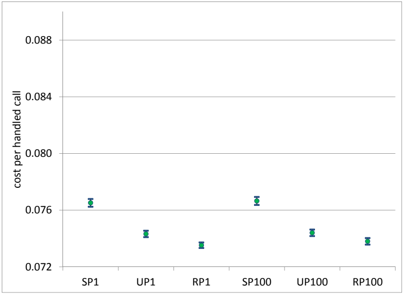
- A network of four large retail-banking call centers in US
  - service rate $\mu = 14.6$ calls/30-min, abandonment rate $\theta = 3.93$ calls/30-min
  - 8am-9pm (13 hours) a day, 5 days a week, schedule updates at 11am

- Shift structure and costs
  - 262 feasible daily schedules (7 and 9-hour shifts, with breaks)
    - cost of 1 per agent per 1/2-hour interval
  - 4,973 potential recourse actions (with 1/2-hour costs)
    - send home (-0.75), overtime (1.5), call in (2.0)

- Arrival data, forecasts, and QoS target
  - last 110 days as testing set
  - forecasts based on previous (rolling) 100 days of data
  - target expected abandonment rate of 3% across scenarios

Updating systematically lowers cost per call



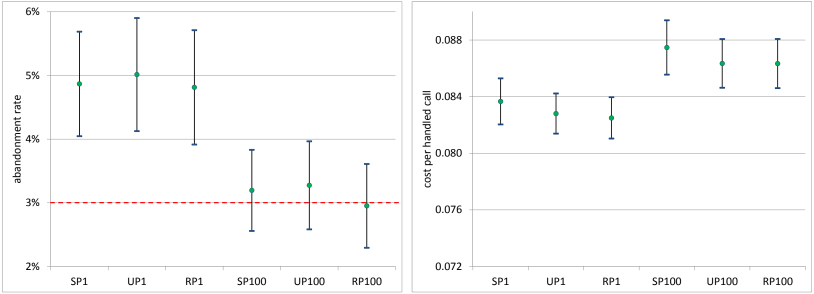


- For abandonment rate: point forecast leads to upward bias
- For cost: pair-wise comparison suggests significant reduction
  - UP100 $\rightarrow$ SP100: average cost 2.9% reduction
  - RP100 $\rightarrow$ SP100: average cost 3.7% reduction

2nd set of empirical tests

- A European retail bank's call center
  - call volume about 15% of the US bank
  - other parameters remain the same
    - schedules, costs
    - service rate, abandonment rate
- Arrival data, forecasts, and QoS target
  - last 76 days as testing set
  - forecasts based on previous (rolling) 100 days of data
  - target expected abandonment rate of 3% across scenarios

Major improvement on abandonment rate



- For abandonment rate: point forecast leads to large upward bias
- For cost: pair-wise comparison suggests significant reduction
  - UP100/RP100 $\rightarrow$ SP100: average cost 1.3% reduction

Summary

- Smooth factor models
- Two-step estimation algorithm
- Within-curve updating
- Combination of Statistics and Operations Research

Curve Registration

July 2016

Goals of FDA

- Represent the data in ways that aid further analysis
  - dimension reduction, functional principal component analysis, ...
- Display the data so as to highlight features
  - smoothing, interpolation, registration, ...
- Study important sources of pattern and variation among the data
  - center, variance, clusters, ...
- Explain variation in a dependent variable by using independent variable information
  - functional linear regression models, functional additive models, ...
- Compare two or more sets of data with respect to certain types of variation
  - curve clustering/classification, functional canonical correlation

Example: girl height curve (RS, 2005)

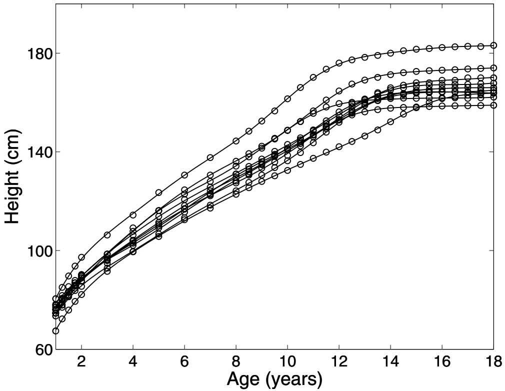
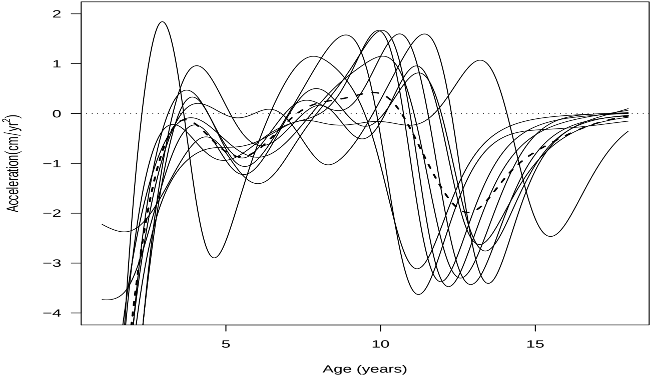
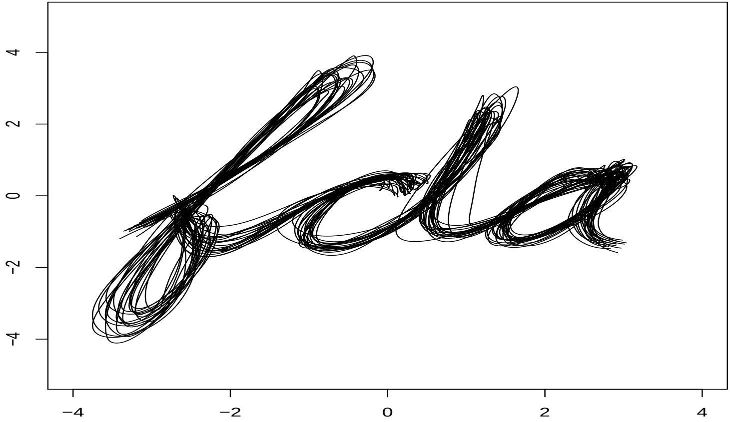


Figure 1.1. The heights of 10 girls measured at 31 ages. The circles indicate the unequally spaced ages of measurement.

Example: girl height acceleration (RS, 2005)



Example: handwriting (RS, 2005)



Example: pinch force (RS, 2005)

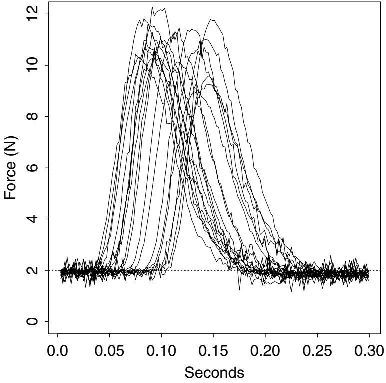
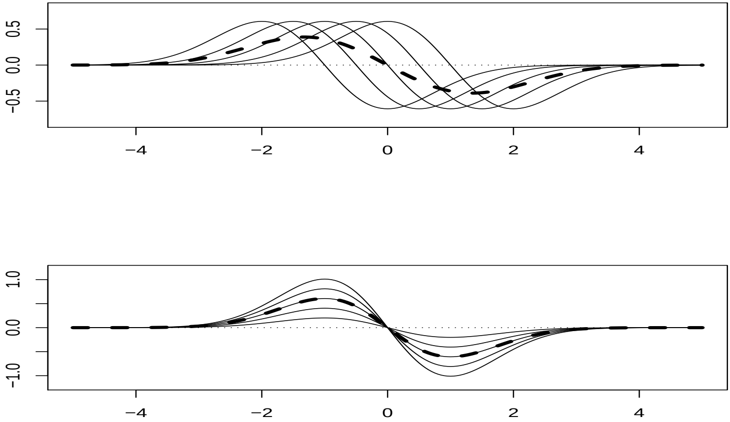


Figure 1.11. Twenty recordings of the force exerted by the thumb and forefinger where a constant background force of two newtons was maintained prior to a brief impulse targeted to reach 10 newtons. Force was sampled 500 times per second.

Example: phase and amplitude variation (RS, 2008)



Registration approaches

- Shift registration
- Landmark or feature registration
- Continuous registration

Shift registration

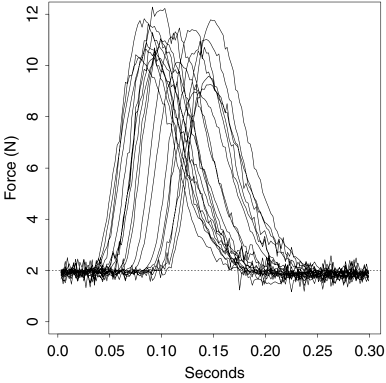


Figure 1.11. Twenty recordings of the force exerted by the thumb and forefinger where a constant background force of two newtons was maintained prior to a brief impulse targeted to reach 10 newtons. Force was sampled 500 times per second.

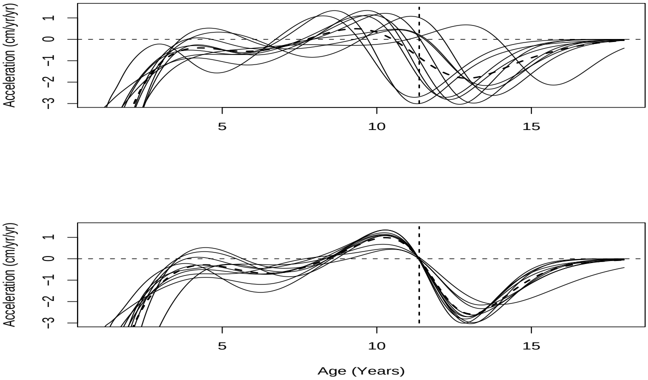
Shift registration

- Registration interval: $\mathcal{T} = [T_1, T_2]$
- Curves: $x_i(t)$ for t in an interval that contains $\mathcal{T}$, $i = 1, \ldots, n$
- Interested in $x_i^*(t) = x_i(t + \delta_i)$ with δ_i being the shift parameter
- Question: how to estimate the δ_i's?

Shift registration - landmark

- Identify a specific feature or landmark for a curve
- Shift each curve so that this feature occurs at a fixed point in time
- Examples:
  - maximum
  - minimum
  - zero-crossing locations

Example: girl height acceleration - landmark registration



Shift landmark registration: undesirable aspects

- location for the feature may be ambiguous
- alignment is local; variations in other regions may be ignored

Shift registration: least squares criterion

- A global registration criterion
- Estimate mean function $\hat{\mu}(t)$ for $t \in \mathcal{T}$
- Define the least squares criterion

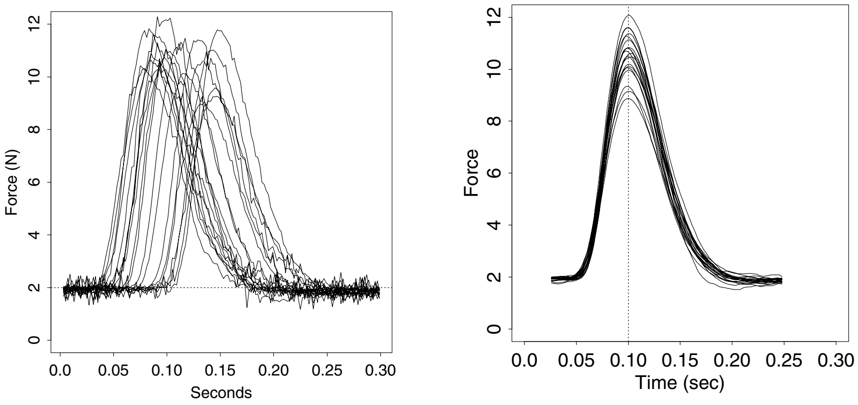
$$\mathrm{REGSSE} = \sum_{i=1}^{n} \int_{\mathcal{T}} [x_i(t + \delta_i) - \hat{\mu}(t)]^2 dt$$

- Integrated squared vertical distances between the shifted curves and the sample mean curve

The Procrustes method

- Update the mean curve $\hat{\mu}(t)$ after the shift registration
- Re-estimate the shift parameters δ_i by minimizing REGSSE
- Iterate until convergence

Shift registration: pinch force



REGSSE minimization: computing issues

- the Newton-Raphson algorithm
- For the pinch force data:
  - align the smoothed curves so that the maximum occurs at 0.1 second
  - get initial estimates of the shifts
  - get initial estimate of the mean curve
  - carry out one step of the Newton-Raphson updating

Landmark registration

- For each curve $x_i(t)$, first identify F features (or landmarks) and the argument values $t_{i,f}$, $f = 1, \ldots, F$

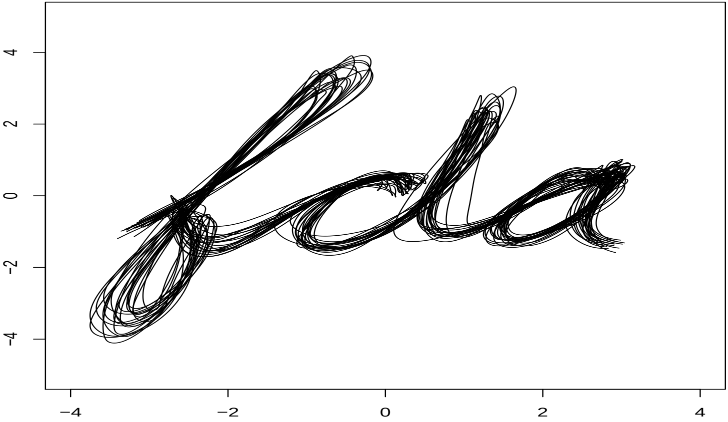
- Goal: construct a transformation h_i for each curve such that the registered curves

$$x_i^*(t) = x_i[h_i(t)]$$

  have similar argument values for any given landmark

- Example: the “fda" handwriting example

Example: handwriting



Example: handwriting

- Normalize the writing time to be between 0 and 2.3 seconds
- Record the position of the pen at a sampling rate of 600 times per second
- I.e., coordinates at the same 1401 equally-spaced time-values
- Data: $x_i(t) = (\mathrm{ScriptX}_i(t), \mathrm{ScriptY}_i(t))$
- Average length of the acceleration vector:

$$\sqrt{(D^2\mathrm{ScriptX}_i)^2 + (D^2\mathrm{ScriptY}_i)^2}$$

Example: handwriting landmark

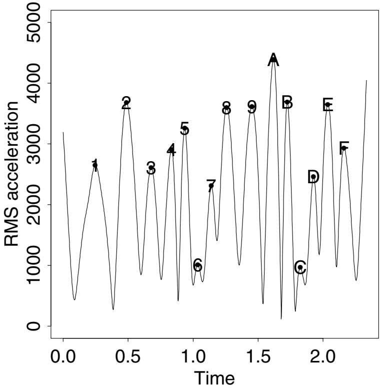


Figure 7.4. The average length of the acceleration vector for the 20 handwriting samples. The characters identify the 15 features used for landmark registration.

Example: handwriting landmark

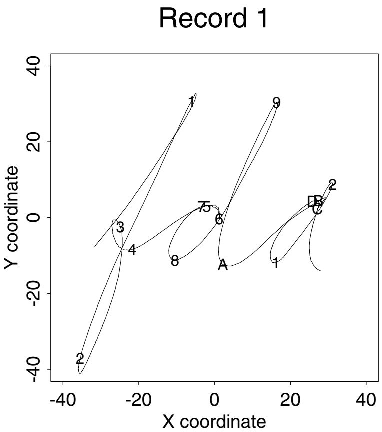


Figure 7.5. The first handwriting curve with the location of the 15 landmarks indicated by the characters used in Figure 7.4.

Example: handwriting landmark timing

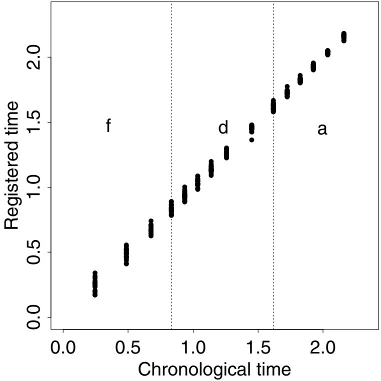


Figure 7.6. The timings of the landmarks for all 20 scripts plotted against the corresponding timings for the mean curve.

Time-warping function

- For each curve, compute a time-warping function $h_i(t)$ such that
  - $h_i(0) = 0$
  - $h_i(2.3) = 2.3$
  - $h_i(t_{0f}) = t_{if}$, $f = 1, \dots, F$:
    - t_{0f}: times of the landmarks on the mean curve
    - t_{if}: times on the ith curve
  - h_i is strictly monotonic
- For this example, use linear interpolations between the points (t_{0f}, t_{if}) to define $h_i(t)$

Example: handwriting landmark warping function for Record 1

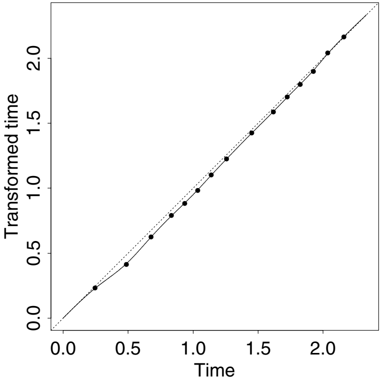


Figure 7.7. The time warping function h_1 estimated for the first record that registers its features with respect to the mean curve.

Continuous registration

- Consider a target curve x_0 and the registered curve $x^* = x(h(t))$
- When the two curves only differ in amplitude not in phase, the points $(x_0(t), x^*(t))$ are on a single line
- Evaluate them at a set of fine grid points of t to form a matrix $\mathbf{X}$
- Then, PCA of $\mathbf{X}^T\mathbf{X}$ should only have one significant eigenvalue
- As a functional analogy, consider

$$\mathbf{T}(h) = \begin{bmatrix} \int x_0^2(t)dt & \int x_0(t)x(h(t))dt \\ \int x_0(t)x(h(t))dt & \int x^2(h(t))dt \end{bmatrix}$$

Continuous registration

- Fitting criterion:

$$\mathrm{MINEIG}(h) = \mu_2(\mathbf{T}(h)),$$

  where μ_2 is the second eigenvalue

- Ideal registration, $\mathrm{MINEIG}(\mathrm{h}) = 0$

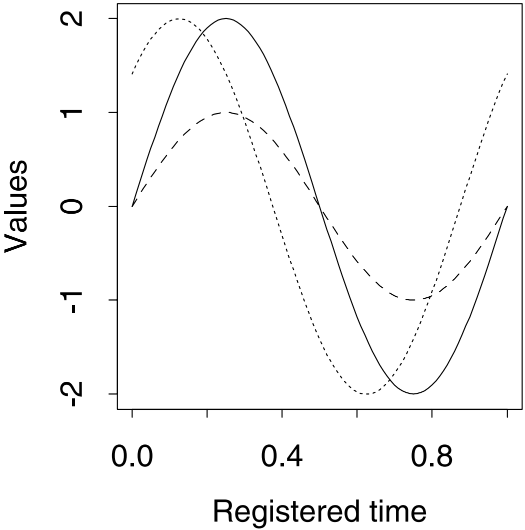
- To impose smoothness on h, extend the criterion to

$$\mathrm{MINEIG}_\lambda(h) = \mathrm{MINEIG}(h) + \lambda \int \{W''(t)\}^2 dt$$

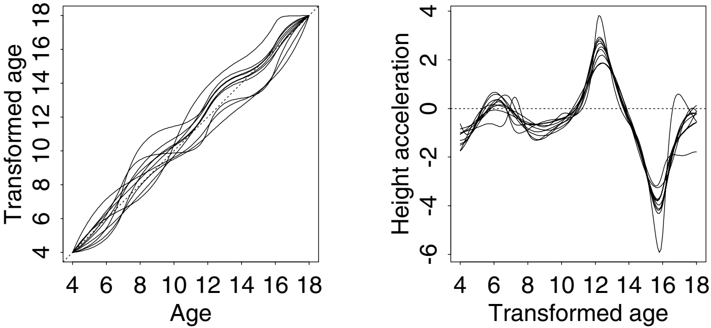
- Minimize either one of the above two criteria

- Basis expansion combined with iterative minimization, see the FDA R/Matlab book

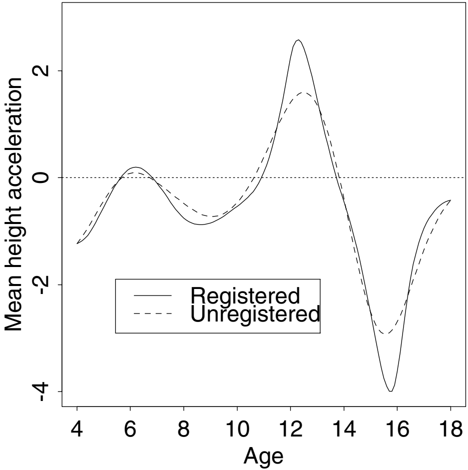
Continuous registration



Example: girl height acceleration - continuous registration



Example: girl height acceleration (RS, 2005)



Example: “Statistical Science" in Chinese