# A Statistical Machine Learning Approach To Yield Curve Forecasting

Rajiv Sambasivan[1] Sourish Das[2]

[1]Department of Computer Science
Chennai Mathematical Institute

[2]Department of Mathematics
Chennai Mathematical Institute

IGIDR 2017

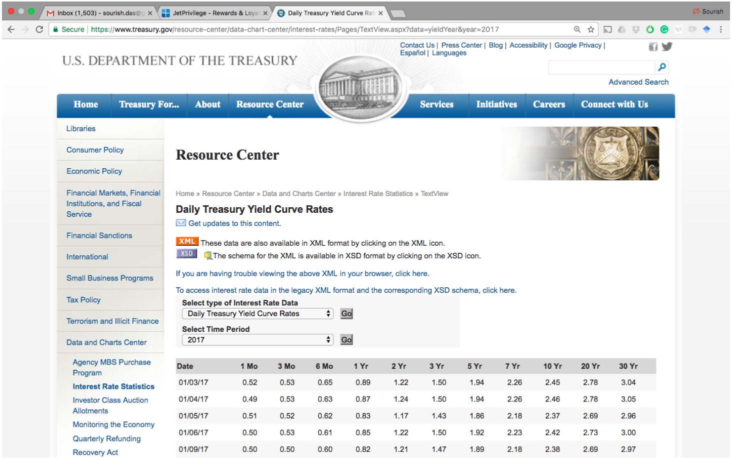
# Outline

- Discussion of problem context - why is estimating the yield curve important?
- Description of Yield Curve Data - what does yield curve data look like?
- Current methods of estimation
- Methodology proposed in this work
- Discussion of Results
- Conclusions

# Problem Context: Yield Curve Estimation

- The bond market is watched closely by financial managers and investors.
- The bond market is a reflector of future economic activity and inflation.
- The future state of the bond market is reflected in the yield curve.
- For this reason estimating the yield curve is very important in finance.

# Description of Yield Curve Data



U.S. DEPARTMENT OF THE TREASURY

Contact Us | Press Center | Blog | Accessibility | Google Privacy | Español | Languages

Advanced Search

Home | Treasury For... | About | Resource Center | Services | Initiatives | Careers | Connect with Us

Libraries
Consumer Policy
Economic Policy
Financial Markets, Financial Institutions, and Fiscal Service
Financial Sanctions
International
Small Business Programs
Tax Policy
Terrorism and Illicit Finance
Data and Charts Center
Agency MBS Purchase Program
Interest Rate Statistics
Investor Class Auction Allotments
Monitoring the Economy
Quarterly Refunding
Recovery Act

## Resource Center

Home » Resource Center » Data and Charts Center » Interest Rate Statistics » TextView

### Daily Treasury Yield Curve Rates

Get updates to this content.

XML These data are also available in XML format by clicking on the XML icon.

XSD The schema for the XML is available in XSD format by clicking on the XSD icon.

If you are having trouble viewing the above XML in your browser, click here.

To access interest rate data in the legacy XML format and the corresponding XSD schema, click here.

Select type of Interest Rate Data: Daily Treasury Yield Curve Rates — Go

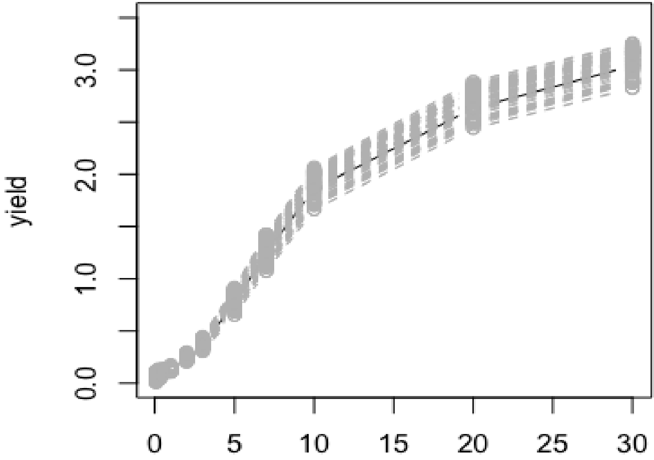
Select Time Period: 2017 — Go

| Date | 1 Mo | 3 Mo | 6 Mo | 1 Yr | 2 Yr | 3 Yr | 5 Yr | 7 Yr | 10 Yr | 20 Yr | 30 Yr |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 01/03/17 | 0.52 | 0.53 | 0.65 | 0.89 | 1.22 | 1.50 | 1.94 | 2.26 | 2.45 | 2.78 | 3.04 |
| 01/04/17 | 0.49 | 0.53 | 0.63 | 0.87 | 1.24 | 1.50 | 1.94 | 2.26 | 2.46 | 2.78 | 3.05 |
| 01/05/17 | 0.51 | 0.52 | 0.62 | 0.83 | 1.17 | 1.43 | 1.86 | 2.18 | 2.37 | 2.69 | 2.96 |
| 01/06/17 | 0.50 | 0.53 | 0.61 | 0.85 | 1.22 | 1.50 | 1.92 | 2.23 | 2.42 | 2.73 | 3.00 |
| 01/09/17 | 0.50 | 0.50 | 0.60 | 0.82 | 1.21 | 1.47 | 1.89 | 2.18 | 2.38 | 2.69 | 2.97 |

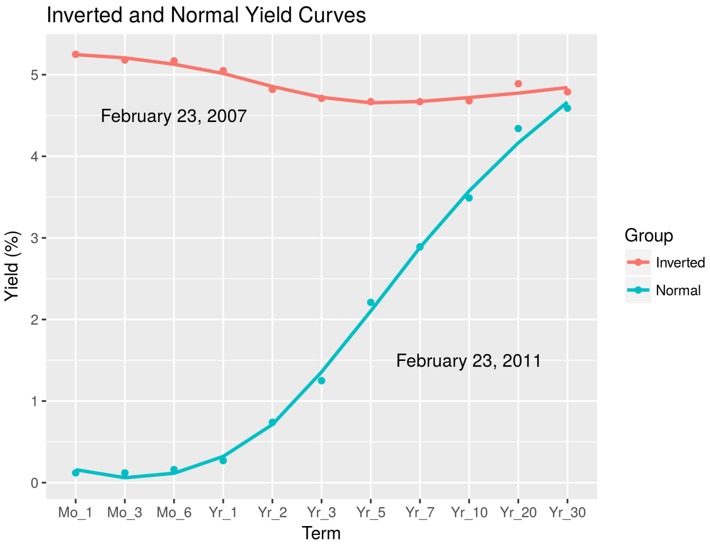
# Description of Yield Curve Data

- Yield curve data (US Treasuries) is available as a time series with a frequency of one day
- A curve is a set of 11 ordered pairs: (term structure, yield for term structure)
- This data can be viewed as a map or a function from term structures to yields.
- This kind of data is called functional data.

# Variability over Yield Curve during 2013



# Normal and Inverted Yield Curves



# Prevalent Methods of Estimating the Yield Curve

- Nelson Siegel Model:
  - Models the yield curve using basis functions:
    $y(\tau) = \beta_1 + \beta_2 \left( \frac{1-e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_3 \left( \frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) + \epsilon(\tau), \quad \epsilon(\tau) \sim N(0, \sigma_\epsilon^2)$
  - To forecast a yield curve, we use the historic values of the $\beta' s$ to forecast future values.
- Multivariate Time Series Model: The yield for each term, $\tau$, is represented in terms of the previous $k$ yields for the same term:

$$\mathbf{y}_i(\tau) = \beta_0 + \beta_1.\mathbf{y}_{i-1}(\tau) + \ldots + \beta_k \mathbf{y}_{i-k}(\tau)$$

  - To forecast the yield curve, we use a standard multi-variate time series package like `vars` in `R`. Future yield estimates are forecast using historic yield estimates.

# Proposed Method - Dynamic Gaussian Process

- Motivation for the proposed method:
  - Gaussian Processes have been applied with great success to model functional data
  - Dynamic Linear Models (Kalman Filter) have been applied with success to model complex time series.
  - We wanted to explore marrying the two approaches.
- Main elements of the approach:
  - A Gaussian Process is used to model the yield, $\mathbf{y}_t$ in terms of the term structure, $\tau$:

$$\mathbf{y}_t = \mu_t(\tau) + \boldsymbol{\epsilon}_t$$

# Dynamic Gaussian Process... continued

- An estimate for the yield at time $t$ is provided using the previous time step $t-1$. The expected value of $\mathbf{y}_t|\mathbf{Y}_{t-1}$ :

$$\begin{aligned}\hat{\mu}_t(\tau*) &= \mathbb{E}(\mu_t(\tau*)|\mathbf{Y}_{t-1}) \\ &= K(\tau*,\tau|\hat{\rho}_{t-1}).[K(\tau,\tau|\hat{\rho}_{t-1}) + \hat{\sigma}^2_{t-1}.\mathbf{I}]^{-1}.\mathbf{y}_{t-1}(\tau).\end{aligned}$$

- Once we have observed the yield curve at time $t$, we can update the posterior process over $\mathbf{y}_t$ as:

$$\begin{aligned}\hat{\mu}_{t.updated}(\tau*) &= \mathbb{E}(\hat{\mu}_t(\tau*)|\mathbf{Y}_t) \\ &= \hat{\mu}_t(\tau*) + K(\tau*,\tau|\hat{\rho}_t).[K(\tau,\tau|\hat{\rho}_t) + \hat{\sigma}^2_t.\mathbf{I}]^{-1}.(\mathbf{y}_t - \hat{\mu}_t(\tau))\end{aligned}$$

# Results

- We compared the performance of the proposed method with the prevalent methods using data over a 10 year period. Results shown below

| Term | GP | MVTS | TSNS |
|---|---|---|---|
| 1 Month | 0.104 | 0.088 | 0.121 |
| 3 Months | 0.071 | 0.066 | 0.080 |
| 6 Months | 0.054 | 0.047 | 0.088 |
| 1 Year | 0.047 | 0.043 | 0.085 |
| 2 Years | 0.052 | 0.055 | 0.088 |
| 3 Years | 0.058 | 0.061 | 0.114 |
| 5 years | 0.065 | 0.068 | 0.126 |
| 7 Years | 0.065 | 0.070 | 0.149 |
| 10 Years | 0.063 | 0.067 | 0.197 |
| 20 Years | 0.061 | 0.065 | 0.977 |
| 30 Years | 0.060 | 0.063 | 10.838 |

Table: RMSE for term structures for all methods

# Discussion of Results

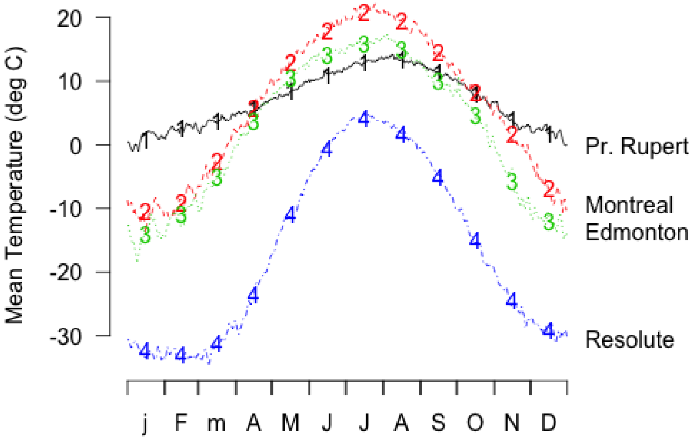
- The multi-variate time series method performs well in the short term structure regions of the yield curve (term structures of one year or less)
- The proposed method does well in the medium and long term regions of the yield curve.
- The proposed method directly yields uncertainty estimates.
- Uncertainty estimates and accuracy in the longer term regions are very useful to analysts and in practice.

# Conclusions

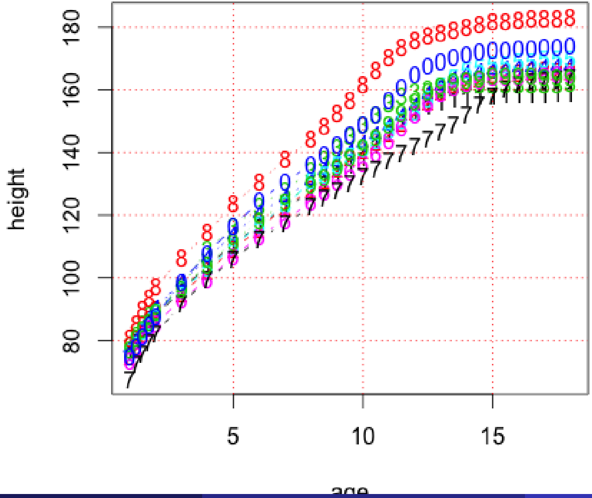
- The proposed method is effective in modeling yield curve data.
- Functional data presents as time series in many domains - hourly electrical load (electrical utilities), hourly requests received by a data center or an application server.
- The effectiveness of this method suggests that this approach could be useful in other application areas as well.

- **More Detail?**
- Functional Data Analysis
- Let's look into some different kind of examples

# Canadian average annual weather cycle



# Berkeley Growth Study



# Cursive Handwriting Samples

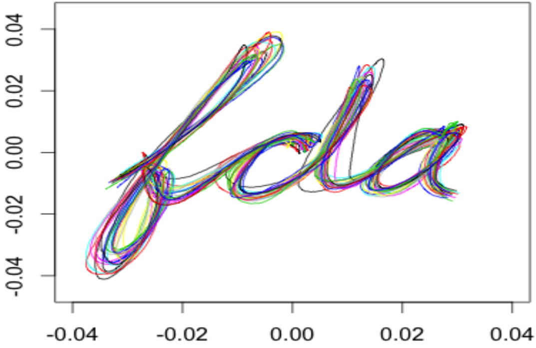
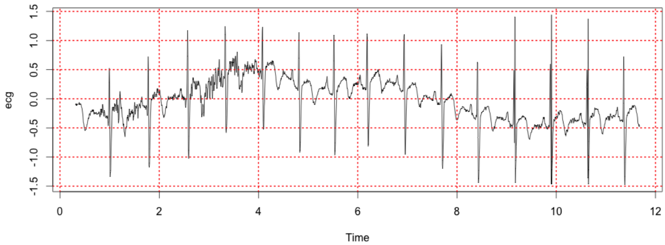
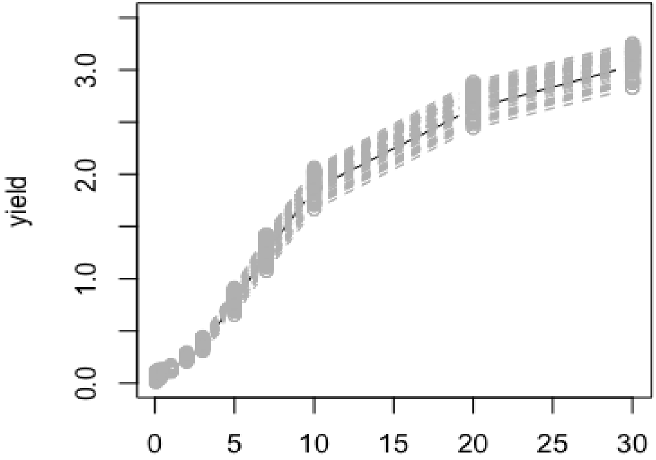



Figure: Measures of position of nib of a pen writing “fda”. 20 replications, measurements taken at 200 hertz.

# ECG Data



# Variability over Yield Curve during 2013



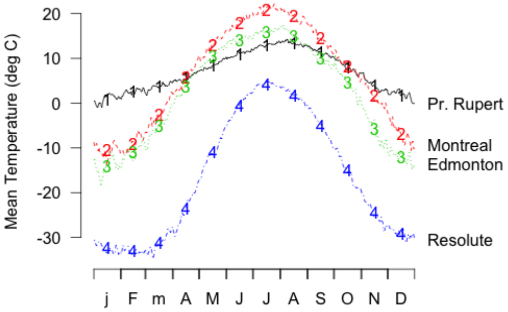
# Features of Functional Data

- Key Feature is smoothness.

$$y_i = f(t_i) + \epsilon_i$$

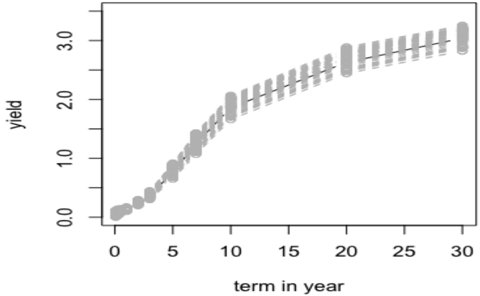
with $t$ is continuum (usually time) and $f(t)$ is smooth function

# Canadian average annual weather cycle




- Model:
  $\mathtt{Temp}_i(t) \approx \beta_{i1} + \beta_{i2}\sin(\pi t/6) + \beta_{i3}\cos(\pi t/6)$,
  where $\mathtt{Temp}_i$ is the temperature function for the $i^{th}$ weather station

# Yield Curve




- Model:

$$y_i(\tau) = \beta_{i1} + \beta_{i2}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{i3}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + \epsilon_i(\tau), \quad \epsilon_i(\tau) \sim N(0, \sigma_\epsilon^2)$$

where $y_i(\tau)$ is the yield curve on $i^{th}$ day

## Basis Expansion

# Representing Functions with Basis Functions

- Consider $i^{th}$ record

$$y_i = f(t_i) + \epsilon_i$$

represents $f(t)$ as

$$f(t) = \sum_{j=1}^{K} \beta_j \phi_j(t) = \boldsymbol{\phi\beta}$$

we say $\phi$ is a basis system for $f(t)$.

# Representing Functions with Basis Functions

- Terms for curvature in linear regression

$$y_i = \beta_1 + \beta_2 t_i + \beta_3 t_i^2 + \ldots + \epsilon_i$$

implies

$$f(t) = \beta_1 + \beta_2 t + \beta_3 t^2 + \ldots$$

# Fourier Basis

- sine cosine functions of incresing frequencies

$$y_i = \beta_1 + \beta_2 \sin(\omega t) + \beta_3 \cos(\omega t) + \beta_4 \sin(2\omega t) + \beta_5 \cos(2\omega t) \ldots + \epsilon_i$$

- constant $\omega = 2\pi/P$ defines the period P of oscillation of the first sine/cosine pair.

- $\phi = \{1, \sin(\omega t), \cos(\omega t), \sin(2\omega t), \cos(2\omega t)...\}$
- $\beta^T = \{\beta_1, \beta_2, \beta_3, \ldots\}$

$$y = \boldsymbol{\phi\beta} + \epsilon$$

# Other Basis

- **Exponential Basis** $\phi = \{1, e^{\lambda_1 t}, e^{\lambda_2 t} ...\}$
- **Gaussian Basis** $\phi = \{1, \exp(-\lambda(t_1 - c)^2), \exp(-\lambda(t_2 - c)^2) ...\}$
- **Basis corresponds to Spline Regression**

$$y = \beta_0 + \sum_{k=1}^{K} \beta_k (t - \xi_k)_+^D + \ldots \epsilon$$

$\phi = \{1, (t - \xi_1)_+^D, (t - \xi_2)_+^D ...\}$

- **Yield Curve - NS Model**: $\phi = \{1, \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right), \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right)\}$, where $K = 3$

# Estimation/Learning

- We are writing the function with its basis expansion

$$y = \phi\boldsymbol{\beta} + \epsilon$$

- Lets assume basis $\phi$ is fully known
- Problem is $\boldsymbol{\beta}$ is unknown - hence we estimate $\boldsymbol{\beta}$.

# Estimation/Learning

- Ordinary Least Square Methods
- Penalized Least Square Methods
- Bayesian Methods

# Penalized Least Square

- The least square criterion with penalty:

$$\text{PSSE} = (y - \phi\beta)^T (y - \phi\beta) + \lambda P(f)$$

  $P(f)$ measures the "roughness" of the $f$
  $\lambda$ represents a continuous tuning parameter

  - $\lambda \uparrow \infty$ roughness increasingly penalized; $f(t)$ becomes smooth
  - $\lambda \downarrow 0$ penalty reduces; $f(t)$ model small shocks and tends to overfit as it move towards OLS

- Essentially $P(f)$ measures the curvature of $f(t)$

# The $D$ Operator

- $Df(t) = \frac{\partial}{\partial t} f(t)$ is the instantaneous slope of $f(t)$
- $D^2 f(t) = \frac{\partial^2}{\partial t^2} f(t)$ is the curvature of $f(t)$
- We measure the size of the curvature for all of $f$ by

$$P(f) = \int [D^2 f(t)]^2 dt$$

# The $D$ Operator

- $Df(t) = \frac{\partial}{\partial t} f(t)$ is the instantaneous slope of $f(t)$
- $D^2 f(t) = \frac{\partial^2}{\partial t^2} f(t)$ is the curvature of $f(t)$
- We measure the size of the curvature for all of $f$ by

$$\begin{aligned} P(f) &= \int [D^2 f(t)]^2 dt \\ &= \int \boldsymbol{\beta}^T [D^2 \boldsymbol{\phi}(t)][D^2 \boldsymbol{\phi}(t)]^T \boldsymbol{\beta} dt \end{aligned}$$

# The $D$ Operator

- $Df(t) = \frac{\partial}{\partial t} f(t)$ is the instantaneous slope of $f(t)$
- $D^2 f(t) = \frac{\partial^2}{\partial t^2} f(t)$ is the curvature of $f(t)$
- We measure the size of the curvature for all of $f$ by

$$
\begin{aligned}
P(f) &= \int [D^2 f(t)]^2 dt \\
&= \int \boldsymbol{\beta}^T [D^2 \boldsymbol{\phi}(t)][D^2 \boldsymbol{\phi}(t)]^T \boldsymbol{\beta} dt \\
&= \boldsymbol{\beta}^T R_2 \boldsymbol{\beta},
\end{aligned}
$$

where $[R_2]_{jk} = \int [D^2 \phi_j(t)][D^2 \phi_k(t)]^T dt$ is the penalty matrix

# Bayesian method

- Model:

$$\mathbf{y} = f(t) + \epsilon$$

  - $\epsilon \sim \mathbf{N}(0, \sigma^2\mathbf{I}) \implies y \sim \mathbf{N}(f(t), \sigma^2\mathbf{I})$

$$f(t) = \phi\boldsymbol{\beta} = \sum_{k=1}^{\infty} \phi_k(t)\beta_k$$

,

- $\boldsymbol{\beta}$ is unknown and want to estimate

  Assuming $\beta$'s are uncorrelated random variable and $\phi_k(t)$ are known deterministic real-valued functions.
- Then due to **Kosambi-Karhunen-Loeve** theorem, we can say that $f(t)$ is a stochastic process.

Gaussian Process Prior

# Gaussian Process Prior

- As $f(t)$ is a stochastic process if we assume $\boldsymbol{\beta} \sim \mathbf{N}(0, \sigma^2\mathbf{I})$ then $f(t) = \phi\boldsymbol{\beta}$ follow Gaussian process.

- Since $f(t)$ is unknown function; therefore induced process on $f(t)$ is known as '**Gaussian Process Prior**'.

  Prior on $\boldsymbol{\beta}$:

$$p(\beta) \propto \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{\beta}^T\boldsymbol{\beta}\right)$$

  Induced Prior on $f = \phi\boldsymbol{\beta}$:

$$p(f) \propto \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{\beta}^T\boldsymbol{\phi}^T\mathbf{K}^{-1}\phi\boldsymbol{\beta}\right)$$

# Gaussian Process Prior

- The prior mean and covariance of $f$ are given by

$$\mathbf{E}[f] = \phi E[\boldsymbol{\beta}] = \phi \boldsymbol{\beta}_0$$

$$\mathbf{cov}[f] = \mathbf{E}[f.f^T] = \phi.\mathbf{E}[\beta.\beta^{\mathbf{T}}]\phi^T = \sigma^2 \phi.\phi^T = \mathbf{K}$$

$$f \quad \sim \quad \mathbf{N}_p(\phi \boldsymbol{\beta}_0, \mathbf{K}), \quad \epsilon \sim \mathbf{N}_p(0, \sigma^2 \mathbf{I})$$

$$\mathbf{y} \quad \sim \quad \mathbf{N}_p(\phi \boldsymbol{\beta}_0, \mathbf{K} + \sigma^2 \mathbf{I})$$

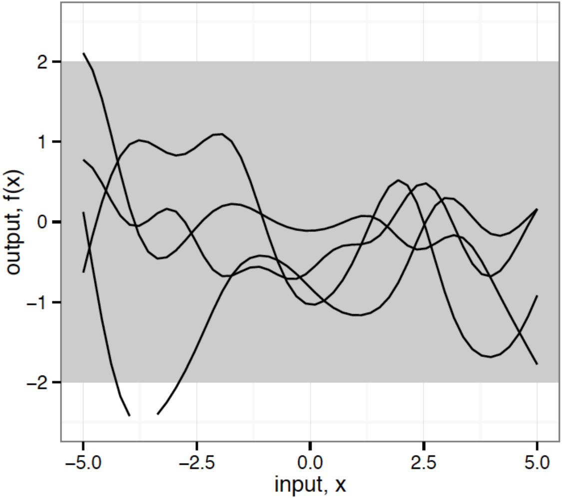
- The estimated value of $y$ for a given $t$ is the mean (expected) value of the functions sampled from from the posterior at that value of $t$.
- The expected value of the estimate at a given $t$ is given by

$$\hat{f}_*(t) = E(f_*|t, Y) = K(t_*, t).[K(t, t) + \sigma^2.\mathbf{I}]^{-1}.\mathbf{y}$$

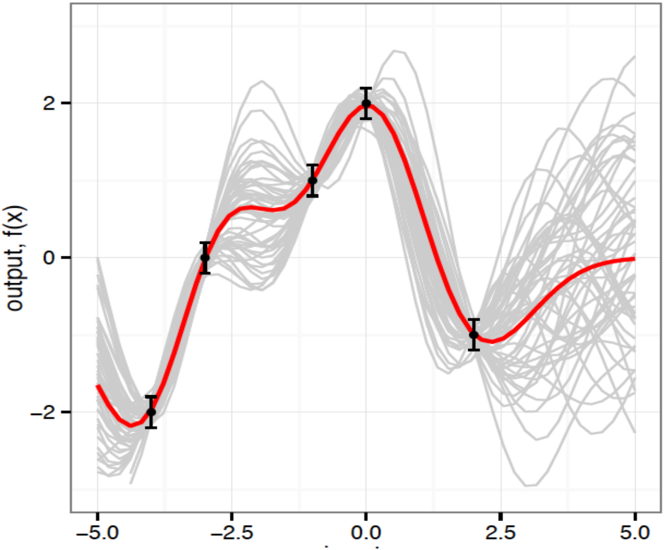
- The variance of the estimate at a given $t$ is given by

$$cov(f_*) = K(t_*, t_*) - K(t_*, t).[K(t, t) + \sigma^2.\mathbf{I}]^{-1}.K(t, t_*)$$

# Gaussian Process Prior



# Estimated Curve using GP Prior



# Dynamic Gaussian Process

- We have used the Kalman Filter Structure
- An estimate for the yield at time $t$ is provided using the previous time step $t-1$. The expected value of $\mathbf{y}_t|\mathbf{Y}_{t-1}$ :

$$\begin{aligned}\hat{\mu}_t(\tau*) &= \mathbb{E}(\mu_t(\tau*)|\mathbf{Y}_{t-1}) \\ &= K(\tau*,\tau|\hat{\rho}_{t-1}).[K(\tau,\tau|\hat{\rho}_{t-1})+\hat{\sigma}^2_{t-1}.\mathbf{I}]^{-1}.\mathbf{y}_{t-1}(\tau).\end{aligned}$$

- Once we have observed the yield curve at time $t$, we can update the posterior process over $\mathbf{y}_t$ as:

$$\begin{aligned}\hat{\mu}_{t.updated}(\tau*) &= \mathbb{E}(\hat{\mu}_t(\tau*)|\mathbf{Y}_t) \\ &= \hat{\mu}_t(\tau*)+K(\tau*,\tau|\hat{\rho}_t).[K(\tau,\tau|\hat{\rho}_t)+\hat{\sigma}^2_t.\mathbf{I}]^{-1}.(\mathbf{y}_t-\hat{\mu}_t(\tau))\end{aligned}$$

**Thank You**

sourish@cmi.ac.in