

The Clean Thesis Style

Ricardo Langner

August 26, 2015
Version: My First Draft

Clean Thesis Style University

Clean**Thesis**

Department of Clean Thesis Style

Institut for Clean Thesis Dev

Clean Thesis Group (CTG)

Documentation

The Clean Thesis Style

Ricardo Langner

- | | |
|--------------------|--|
| <i>1. Reviewer</i> | Jane Doe
Department of Clean Thesis Style
Clean Thesis Style University |
| <i>2. Reviewer</i> | John Doe
Department of Clean Thesis Style
Clean Thesis Style University |
| <i>Supervisors</i> | Jane Doe and John Smith |

August 26, 2015

Ricardo Langner

The Clean Thesis Style

Documentation, August 26, 2015

Reviewers: Jane Doe and John Doe

Supervisors: Jane Doe and John Smith

Clean Thesis Style University

Clean Thesis Group (CTG)

Institut for Clean Thesis Dev

Department of Clean Thesis Style

Street address

Postal Code and City

Abstract

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Abstract (different language)

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Acknowledgement

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Contents

1	Introduction	1
1.1	Postcards: My Address	2
1.2	Motivation and Problem Statement	2
1.3	Results	3
1.3.1	Some References	3
1.4	Thesis Structure	3
2	Forecasting discount curves with Kernel Regularized Least Squares	7
2.1	Introduction	7
2.2	Sensitivity of the response to a change in the covariates	11
2.3	Description of DNS-KRLS model	13
2.4	Description of the KRLS model applied to observed dates and time to maturities	15
2.5	Numerical examples	16
2.6	Conclusion	16
3	Conclusion	21
3.1	System Section 1	21
3.2	System Section 2	22
3.3	Future Work	23
	Bibliography	25

Introduction

“You can’t do better design with a computer, but you can speed up your work enormously.

— Wim Crouwel

(Graphic designer and typographer)

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A

blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

1.1 Postcards: My Address

Ricardo Langner

Alfred-Schrapel-Str. 7

01307 Dresden

Germany

1.2 Motivation and Problem Statement

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of

the original language. There is no need for special content, but the length of words should match the language.

[Jür00; Jür95; Mie11a; KJUM11; App10a; App10b; App10c]

1.3 Results

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

1.3.1 Some References

[@Fre10; @Mie11b]

1.4 Thesis Structure

Chapter ??

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Chapter ??

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Chapter ??

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Chapter ??

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Chapter 3

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Forecasting discount curves with Kernel Regularized Least Squares

2.1 Introduction

In this chapter, we apply Kernel Regularized Least Squares (KRLS) learning methods to Yield Curve forecasting. By 'Yield Curve', we actually mean a discount curves. That is, we consider that the curves used in the examples do not include counterparty credit risk, and we focus on the forecasting problem. Two types of formulations of the forecasting problem are tested. One relying on the popular Dynamic Nelson-Siegel (DNS) framework from [DL06], and another one, in which we apply the KRLS directly to the Yield Curve observation dates and time to maturities, to model the spot rates.

In the DNS framework [DL06], each cross-section of yields observed over time is fitted by using the Nelson-Siegel [NS87] model. This cross-section fitting produces three time series of parameters (more details in the next section) representing the evolution of the level, slope, and curvature of the Yield Curve. The KRLS model is applied to forecasting the time series of parameters, using a technique which is similar to the one described in [Ext+16]. And to finish, the forecast obtained for the trivariate time series are plugged into the Nelson-Siegel model, to deduce forecast for the cross-sections of yields.

The second approach is a machine learning/data-driven one, in which we put no specific constraint on the model to reproduce the specific Yield Curve stylized facts. The regularization parameters inherent to the KRLS model will act as implicit constraints, that cause the model to converge as close as possible to reproducing these stylized facts. In this latter approach, we are mostly interested in the model with the *best* out-of-sample error. As a consequence, the technique as is, is probably less adapted than the former framework based on DNS (in its arbitrage-free version) to no-arbitrage pricing (if it's required).

To introduce KRLS, we start by describing the ridge regression [HK70] and the *kernel trick* applied to ridge regression. Then, we make a link between this kernel ridge

regression and KRLS. In a ridge regression setting, we want to explain an observed variable $y \in \mathbb{R}^n$, as a linear function of p predictors stored in a matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$. For $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, p\}$ we have:

$$X_{ij} =: \mathbf{x}_i^{(j)}$$

We will denote the i^{th} row of \mathbf{X} as \mathbf{x}_i , and its j^{th} column as $\mathbf{x}^{(j)}$. Hence, we are searching for the parameters $\beta = (\beta_1, \dots, \beta_p)^T$ verifying:

$$\text{ArgMin}_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \left(y_i - \mathbf{x}_i^T \beta \right)^2$$

under the constraint

$$\|\beta\|_2^2 \leq s$$

The solution to this problem is given directly by the formula:

$$\hat{\beta} = \left(\mathbf{X}^T \mathbf{X} + \lambda I_{p \times p} \right)^{-1} \mathbf{X}^T y$$

where λ is a Lagrange multiplier having a unique correspondance with s , and a regularization parameter preventing the model from overfitting the observed data contained in y . In the case where we want to explain y as a function Φ of the predictors, we have a similar expression:

$$\hat{\beta} = \left(\Phi(\mathbf{X})^T \Phi(\mathbf{X}) + \lambda I_{p \times p} \right)^{-1} \Phi(\mathbf{X})^T y$$

where:

$$\Phi(\mathbf{X})_{ij} = \Phi(\mathbf{x}_i^{(j)})$$

Now, by using the Woodbury identity (cite Gene H. Golub and Charles F. van Loan. Matrix Computations and cite Max Welling The Kalman filter, Lecture Note) for \mathbf{P} and \mathbf{R} positive definite

$$\left(\mathbf{P}^{-1} + \mathbf{B}^T \mathbf{R}^{-1} \mathbf{B} \right)^{-1} \mathbf{B}^T \mathbf{R}^{-1} = \mathbf{P} \mathbf{B}^T \left(\mathbf{B} \mathbf{P} \mathbf{B}^T + \mathbf{R} \right)^{-1}$$

The solution to the ridge regression problem can be re-written as:

$$\hat{\beta} = \Phi(\mathbf{X})^T \left(\Phi(\mathbf{X}) \Phi(\mathbf{X})^T + \lambda I_{n \times n} \right)^{-1} y$$

This relationship can be useful in the case where $n \ll p$. That is, when there is a high number of predictors compared to the number of observations (cite Exterkate (2016)). Indeed, with this new relationship, we are no longer inverting a $p \times p$ matrix, but a $n \times n$ matrix. That's the *kernel trick*. And if some new observations arrive, and are stored in \mathbf{X}^* , the new values predicted by the model will be given by:

$$y^* = \Phi(\mathbf{X}^*)\hat{\beta} = \Phi(\mathbf{X}^*)\Phi(\mathbf{X})^T \left(\Phi(\mathbf{X})\Phi(\mathbf{X})^T + \lambda I_{n \times n} \right)^{-1} y$$

Which we re-write as:

$$y^* = \mathbf{K}^* (\mathbf{K} + \lambda I_{n \times n})^{-1} y$$

\mathbf{K} is a *kernel*; the empirical covariance matrix of $\Phi(\mathbf{X})^T$ (modulo a $1/p$ factor), in the case where the rows of $\Phi(\mathbf{X})$ are centered. Now, in the case of KRLS, the problem we are trying to solve is:

$$\text{ArgMin}_{c \in \mathbb{R}^n} \sum_{i=1}^n \left(y_i - K_i^T c \right)^2$$

where K_i is the i^{th} row of \mathbf{K} , with:

$$\mathbf{K}_{ij} =: K(\mathbf{x}_i, \mathbf{x}_j) = f(\|\mathbf{x}_i - \mathbf{x}_j\|_1) \text{ or } f(\|\mathbf{x}_i - \mathbf{x}_j\|_2)$$

As described in (cite Halmueller & Hazlett (2013)), two approaches can be used to interpret/motivate the KRLS: a similarity-based view and the superposition of Gaussians view. Here, we refer only to the first one, which is the most intuitive to us, and is also the one described in Ferwerda, Hainmueller et Hazlett (2017). Indeed, here, the i^{th} observation y_i is explained as a linear combination of functions measuring the similarity/dissimilarity between its characteristics gathered in \mathbf{x}_i , and the other observations from the training set, \mathbf{x}_j , $j \neq i$:

$$y_i = \sum_{j=1}^n c_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{2.1}$$

But again, to prevent the model from overfitting the observed data and not being able to generalize, we need to constrain the parameters c under a certain norm defined by the matrix \mathbf{K} (cite Hofmann, Schoelkopf, Smola (2008)):

$$\|c\|_K^2 = c^T \mathbf{K} c \leq s$$

The solution to this new problem is:

$$\hat{c} = (\mathbf{K} + \lambda I_{n \times n})^{-1} \mathbf{y}$$

where λ is a Lagrange multiplier having a unique correspondance with s , and a regularization parameter. And for new observations arriving for the model, we have a solution which is identical to the one that we had for kernel ridge regression:

$$y^* = \mathbf{K}^* (\mathbf{K} + \lambda I_{n \times n})^{-1} y$$

Many other types of kernels could be envisaged for \mathbf{K} , allowing to take into account nonlinearities and the various complexities of the covariance structure. One of the most popular kernels is the Gaussian kernel, also called *squared exponential* kernel, defined for $i < j$ by:

$$K_{Gauss}(\mathbf{x}_i, \mathbf{x}_j) = \sigma^2 \exp \left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2l^2} \right)$$

where l is a characteristic length-scale controlling the distance between peaks of the covariance function. σ^2 is the marginal variance, obtained when $\mathbf{x}_i = \mathbf{x}_j$. Both l , σ^2 are used as the machine learning model's hyperparameters, along with the regularization parameter λ .

This kernel is however often judged as being too smooth for most typical optimization problems (cite Rasmussen et al.). Some other kernels that could be interesting for machine learning (cite Rasmussen et al.) belong to the Matérn class of covariance functions. If we define $r := \|\mathbf{x}_i - \mathbf{x}_j\|_2$, the most used for machine learning problems (cite Rasmussen et al.) are:

$$\mathbf{K}_{ij} = K_{3/2}(r) = \sigma^2 \left(1 + \frac{\sqrt{3}r}{l} \right) \exp \left(-\frac{\sqrt{3}r}{l} \right)$$

and

$$\mathbf{K}_{ij} = K_{5/2}(r) = \sigma^2 \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2} \right) \exp \left(-\frac{\sqrt{5}r}{l} \right)$$

Another interesting feature of KRLS learning, is the possibility to derive estimators for the marginal effects of the covariates on the response. For example, (and like in Ferwerda, Hainmueller et Hazlett (2017)), since we have the relationship 2.1, we can write for a fixed $j_0 \in \{1, \dots, p\}$ and for any $k \in \{1, \dots, n\}$, :

$$\frac{\partial y_i}{\partial \mathbf{x}_k^{(j_0)}} = \sum_{j=1}^n c_j \frac{dK(\mathbf{x}_i, \mathbf{x}_j)}{d\mathbf{x}_k^{(j_0)}} = c_k \frac{\partial K(\mathbf{x}_i, \mathbf{x}_k)}{\partial \mathbf{x}_k^{(j_0)}} \quad (2.2)$$

That's an approximation of how much of an increase (at the first order) we obtain in y_i , for a slight change in $\mathbf{x}_k^{(j_0)}$. An average marginal effect of the j_0^{th} covariate on the i^{th} observation of the response y can thus be obtained as:

$$\frac{1}{n} \sum_{k=1}^n \frac{\partial y_i}{\partial x_k^{(j_0)}} = \frac{1}{n} \sum_{k=1}^n c_k \frac{\partial K(\mathbf{x}_i, \mathbf{x}_k)}{\partial \mathbf{x}_k^{(j_0)}} \quad (2.3)$$

In section 2.2, we derive formulas for the sensitivities of the response variable y_i to a change in the covariates $\mathbf{x}^{(j)}$. We do this for Gaussian, Matérn 3/2 and Matérn 5/2 kernels. These sensitivities can then be plugged into formula 2.3, in order to obtain average marginal effects of the covariates on the response.

2.2 Sensitivity of the response to a change in the covariates

We let:

$$r^2 := \|\mathbf{x}_i - \mathbf{x}_k\|_2^2 = (\mathbf{x}_i - \mathbf{x}_k)^T (\mathbf{x}_i - \mathbf{x}_k)$$

Hence:

$$\frac{\partial r^2}{\partial \mathbf{x}_k^{(j_0)}} = -2 \left(\mathbf{x}_i^{(j_0)} - \mathbf{x}_k^{(j_0)} \right)$$

And:

$$\frac{\partial r}{\partial \mathbf{x}_k^{(j_0)}} = \frac{1}{2r} \frac{\partial r^2}{\partial \mathbf{x}_k^{(j_0)}} = -\frac{1}{r} \left(\mathbf{x}_i^{(j_0)} - \mathbf{x}_k^{(j_0)} \right)$$

As a consequence:

- For the **Gaussian kernel**

$$(\mathbf{x}_i, \mathbf{x}_k) \mapsto \sigma^2 \exp\left(-\frac{r^2}{2l^2}\right)$$

We have:

$$\begin{aligned} \frac{\partial K_{Gauss}(\mathbf{x}_i, \mathbf{x}_k)}{\partial \mathbf{x}_k^{(j_0)}} &= -\frac{\sigma^2}{2l^2} \exp\left(-\frac{r^2}{2l^2}\right) \frac{\partial r^2}{\partial \mathbf{x}_k^{(j_0)}} \\ &= -\frac{\sigma^2}{2l^2} \exp\left(-\frac{r^2}{2l^2}\right) [-2(\mathbf{x}_i^{(j_0)} - \mathbf{x}_k^{(j_0)})] \\ &= \frac{\sigma^2}{l^2} \exp\left(-\frac{r^2}{2l^2}\right) (\mathbf{x}_i^{(j_0)} - \mathbf{x}_k^{(j_0)}) \\ &= \frac{(\mathbf{x}_i^{(j_0)} - \mathbf{x}_k^{(j_0)})}{l^2} K_{Gauss}(\mathbf{x}_i, \mathbf{x}_k) \end{aligned}$$

- For the **Matérn 3/2 kernel**

$$(\mathbf{x}_i, \mathbf{x}_k) \mapsto \sigma^2 \left(1 + \frac{\sqrt{3}r}{l}\right) \exp\left(-\frac{\sqrt{3}r}{l}\right)$$

We have:

$$\begin{aligned} \frac{\partial K_{3/2}(\mathbf{x}_i, \mathbf{x}_k)}{\partial \mathbf{x}_k^{(j_0)}} &= \sigma^2 \frac{\sqrt{3}}{l} \exp\left(-\frac{\sqrt{3}}{l}r\right) \left[\frac{\partial r}{\partial \mathbf{x}_k^{(j_0)}} - \frac{\partial r}{\partial \mathbf{x}_k^{(j_0)}} \left(1 + \frac{\sqrt{3}}{l}r\right) \right] \\ &= \sigma^2 \frac{\sqrt{3}}{l} \frac{\partial r}{\partial \mathbf{x}_k^{(j_0)}} \exp\left(-\frac{\sqrt{3}}{l}r\right) \left[1 - \left(1 + \frac{\sqrt{3}}{l}r\right) \right] \\ &= \sigma^2 \frac{\sqrt{3}}{l} \exp\left(-\frac{\sqrt{3}}{l}r\right) \frac{\sqrt{3}(\mathbf{x}_i^{(j_0)} - \mathbf{x}_k^{(j_0)})}{l} \end{aligned}$$

- For the **Matérn 5/2 kernel**

$$(\mathbf{x}_i, \mathbf{x}_k) \mapsto \sigma^2 \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2}\right) \exp\left(-\frac{\sqrt{5}r}{l}\right)$$

We have:

$$\begin{aligned}
\frac{\partial K_{5/2}(\mathbf{x}_i, \mathbf{x}_k)}{\partial \mathbf{x}_k^{(j_0)}} &= \sigma^2 \exp\left(-\frac{\sqrt{5}}{l}r\right) \left[\left(\frac{\sqrt{5}}{l} \frac{\partial r}{\partial \mathbf{x}_k^{(j_0)}} + \frac{5}{3l^2} \frac{\partial r^2}{\partial \mathbf{x}_k^{(j_0)}} \right) - \frac{\sqrt{5}}{l} \frac{\partial r}{\partial \mathbf{x}_k^{(j_0)}} \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2} \right) \right] \\
&= \sigma^2 \exp\left(-\frac{\sqrt{5}}{l}r\right) \left[\frac{5}{3l^2} \frac{\partial r^2}{\partial \mathbf{x}_k^{(j_0)}} - \frac{\sqrt{5}}{l} \frac{\partial r}{\partial \mathbf{x}_k^{(j_0)}} \left(\frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2} \right) \right] \\
&= \sigma^2 \exp\left(-\frac{\sqrt{5}}{l}r\right) (\mathbf{x}_i^{(j_0)} - \mathbf{x}_k^{(j_0)}) \left[-2\frac{5}{3l^2} + \frac{1}{r} \frac{\sqrt{5}}{l} \left(\frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2} \right) \right] \\
&= \sigma^2 \exp\left(-\frac{\sqrt{5}}{l}r\right) \frac{5}{3l^2} \left(1 + \frac{\sqrt{5}r}{l} \right) (\mathbf{x}_i^{(j_0)} - \mathbf{x}_k^{(j_0)})
\end{aligned}$$

2.3 Description of DNS-KRLS model

In the DNS framework ([DL06]), the spot interest rates observed at time t , for time to maturity τ are modeled as:

$$R_t(\tau) = \alpha_{1,t} + \alpha_{2,t} \left(\frac{1 - e^{-\tau/\eta}}{e^{-\tau_1/\eta}} \right) + \alpha_{3,t} \left(\frac{1 - e^{-\tau/\eta}}{e^{-\tau_1/\eta}} - e^{-\tau/\eta} \right) \quad (2.4)$$

If these spot interest rates $R_t(\tau)$ are observed at increasing dates $t = t_1 < \dots < t_n$, for increasing time to maturities $\tau = \tau_1 < \dots < \tau_p$, the factor loadings are the vectors of length p :

$$(1, \dots, 1)^T$$

and

$$\left(\frac{1 - e^{-\tau_1/\eta}}{e^{-\tau_1/\eta}}, \dots, \frac{1 - e^{-\tau_p/\eta}}{e^{-\tau_p/\eta}} \right)^T$$

and

$$\left(\frac{1 - e^{-\tau_1/\eta}}{e^{-\tau_1/\eta}} - e^{-\tau_1/\eta}, \dots, \frac{1 - e^{-\tau_p/\eta}}{e^{-\tau_p/\eta}} - e^{-\tau_p/\eta} \right)^T$$

They are used to represent respectively the level, slope, and curvature of the Yield Curve. Estimations of $\alpha_{i,t}$, $i = 1, \dots, 3$ are obtained for each cross-section of yields (that is, for each fixed date t) by taking a fixed η , and doing a least squares regression of the spot rates observed at time t on these factor loadings. The three time series $(\alpha_{i,t})_t$, $i = 1, \dots, 3$ associated to the loadings for each cross-section of yields, are those that we wish to forecast simultaneously, by using KRLS learning.

For doing this, we store the three time series in a response matrix \mathbf{Y} , and their lags in a matrix of predictors \mathbf{X} . We will then have, for new observations stored in \mathbf{X}^* :

$$\mathbf{Y}^* = \mathbf{K}^* (\mathbf{K} + \lambda I_{n \times n})^{-1} \mathbf{Y}$$

We consider $p \in \mathbb{N}^*$ time series $(X_t^{(j)})_{t \geq 0}, j = 1, \dots, p$, observed at $n \in \mathbb{N}^*$ discrete dates. We are interested in obtaining simultaneous forecasts of the p time series at time $n + h, h \in \mathbb{N}^*$, by allowing each of the p variables to be influenced by the others (in the spirit of VAR models, see [Lüt05]).

For this purpose, we use $k < n$ lags of each of the observed p time series. The output variables to be explained are:

$$Y^{(j)} = \left(X_n^{(j)}, \dots, X_{k+1}^{(j)} \right)^T \quad (2.5)$$

for $j \in \{1, \dots, p\}$. Where $X_n^{(j)}$ is the most contemporaneously observed value of the j^{th} time series, and $X_{k+1}^{(j)}$ was observed k dates earlier in time for $(X_t^{(j)})_{t \geq 0}$. These output variables are stored in:

$$\mathbf{Y} \in \mathbb{R}^{(n-k) \times p}$$

and the predictors are stored in:

$$\mathbf{X} \in \mathbb{R}^{(n-k) \times (k \times p)}$$

where \mathbf{X} consists in p blocks of k lags, for each one of the observed p time series. For example, the j_0^{th} block of \mathbf{X} , for $j_0 \in \{1, \dots, p\}$ contains in columns:

$$\left(X_{n-i}^{(j_0)}, \dots, X_{k+1-i}^{(j_0)} \right)^T \quad (2.6)$$

with $i \in \{1, \dots, k\}$. If we consider $p = 2$ time series $(X_{t_1}^{(1)}, \dots, X_{t_5}^{(1)})$ and $(X_{t_1}^{(2)}, \dots, X_{t_5}^{(2)})$ observed at $n = 5$ dates $t_1 < \dots < t_5$, with $k = 2$ lags, the response variables are stored in:

$$\mathbf{Y} = \begin{pmatrix} X_{t_5}^{(1)} & X_{t_5}^{(2)} \\ X_{t_4}^{(1)} & X_{t_4}^{(2)} \\ X_{t_3}^{(1)} & X_{t_3}^{(2)} \end{pmatrix}$$

The predictors are stored in:

$$\mathbf{X} = \begin{pmatrix} X_{t_4}^{(1)} & X_{t_3}^{(1)} & X_{t_4}^{(2)} & X_{t_3}^{(2)} \\ X_{t_3}^{(1)} & X_{t_2}^{(1)} & X_{t_3}^{(2)} & X_{t_2}^{(2)} \\ X_{t_2}^{(1)} & X_{t_1}^{(1)} & X_{t_2}^{(2)} & X_{t_1}^{(2)} \end{pmatrix}$$

It is also possible to add other regressors, such as dummy variables, or indicators of special events. In this situation, as discussed in [Ext+16], we can avoid the constraining of these dummy variables, in a Kernel ridge regression with unpenalized terms. Here, we consider only the inclusion of the observed time series' lags in the model.

2.4 Description of the KRLS model applied to observed dates and time to maturities

In this other framework, we consider that the response variable is the spot interest rate observed at time t , for time to maturity τ , $R(t, \tau)$. The predictors are the observation date, and the time to maturity. In this setting, we use the following weighted distance between the vectors (t, τ) in the Gaussian, Matérn 3/2 and Matérn 5/2 kernels:

$$r = \sqrt{\frac{\|t_i - t_j\|_2^2}{\theta_1^2} + \frac{\|\tau_i - \tau_j\|_2^2}{\theta_2^2}}$$

So that here, the spot rates values are explained as linear combination of distances between vectors of time to maturities and observation dates (t_i, τ_i) and (t_j, τ_j) . In this setting, if we consider 10 spot rates observed at 2 dates $t_1 < t_2$ and 5 time to maturities τ_1, \dots, τ_5 , the response variable is:

$$\mathbf{Y} = (R(t_1, \tau_1), \dots, R(t_1, \tau_5), R(t_2, \tau_1), \dots, R(t_2, \tau_5))^T$$

and the predictors are

$$\mathbf{X} = \begin{pmatrix} \tau_1 & t_1 \\ \vdots & \vdots \\ \tau_5 & t_1 \\ \tau_1 & t_2 \\ \vdots & \vdots \\ \tau_5 & t_2 \end{pmatrix}$$

Tab. 2.1: Summary of observed discount rates from Deutsche Bundesbank website, from 2002 to the end 2015

Maturity	Min	1st Qrt	Median	3rd Qrt	Max
1	-0.116	0.858	2.045	3.072	5.356
5	0.170	1.327	2.863	3.807	5.146
15	0.711	2.616	3.954	4.702	5.758
30	0.805	2.594	3.962	4.814	5.784
50	0.749	2.647	3.630	4.590	5.467

If some new observations arrive at time t_3 in the model, these will be stored in:

$$\mathbf{X}^* = \begin{pmatrix} \tau_1 & t_3 \\ \vdots & \vdots \\ \tau_5 & t_3 \end{pmatrix}$$

2.5 Numerical examples

The following examples are not exhaustive benchmarks, but aim at illustrating the forecasting capabilities of the models described in this paper. We use calibrated discount rates data from Deutsche Bundesbank website, observed on a monthly basis, from the beginning of 2002 to the end 2015. There are 167 curves, observed at 50 maturities in the dataset. Only 15 time to maturities are used in these examples: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 25, 30.

In figure 2.1, we present the data that we use, and table 2.1 contains a summary of these data; the minimum, maximum, median, first and third quartiles of the discount rates observed at given maturities. There are alternate cycles of increases and decreases of the discount rates, with generally a decreasing trend. Some of the discount rates, at the most recent dates, and lower maturities, are negative.

We use the same grids for all the models, in order to ease the comparability, and to avoid too much manual tuning for the optimal set of parameters. For both types of models, we consider 5 values of σ , l and λ (kernels' parameters and regularization parameters) comprised between and for the DNS-KRLS model, we use 5 values η comprised between the minimum observed time to maturity and the maximum observed time to maturity.

2.6 Conclusion

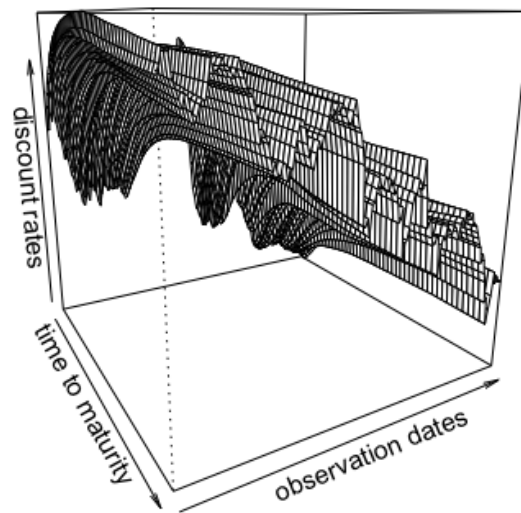


Fig. 2.1: Observed discount rates from Deutsche Bundesbank website, from 2002 to the end 2015

Conclusion

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

3.1 System Section 1

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the

alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

3.2 System Section 2

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference

between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

And after the second paragraph follows the third paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

3.3 Future Work

After this fourth paragraph, we start a new paragraph sequence. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words

should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Bibliography

- [Ber+15] Christoph Bergmeir, Rob J Hyndman, Bonsoo Koo, et al. „A Note on the Validity of Cross-Validation for Evaluating Time Series Prediction“. In: *Monash University, Department of Econometrics and Business Statistics, Tech. Rep.* (2015).
- [Bon+15] F Bonnin, F Combes, F Planchet, and M Tammar. „Un modèle de projection pour des contrats de retraite dans le cadre de l'ORSA“. In: *Bulletin Français d'Actuariat* 14 (2015), pp. 107–129.
- [BT97] Phelim P Boyle and Ken Seng Tan. „Quasi-Monte Carlo methods“. In: *International AFIR Colloquium Proceedings, Australia*. Vol. 1. 1997, pp. 1–24.
- [Car98] Rich Caruana. „Multitask learning“. In: *Learning to learn*. Springer, 1998, pp. 95–133.
- [Cha+92] Kanad Chakraborty, Kishan Mehrotra, Chilukuri K Mohan, and Sanjay Ranka. „Forecasting the behavior of multivariate time series using neural networks“. In: *Neural networks* 5.6 (1992), pp. 961–970.
- [Cor09] Thomas H Cormen. *Introduction to algorithms*. MIT press, 2009.
- [DC10] Satchidananda Dehuri and Sung-Bae Cho. „A comprehensive survey on functional link neural networks and an adaptive PSO–BP learning for CFLNN“. In: *Neural Computing and Applications* 19.2 (2010), pp. 187–205.
- [DL06] Francis X Diebold and Canlin Li. „Forecasting the term structure of government bond yields“. In: *Journal of econometrics* 130.2 (2006), pp. 337–364 (cit. on pp. 7, 13).
- [DR13] Francis X Diebold and Glenn D Rudebusch. *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach*. Princeton University Press, 2013.
- [DS15] Christophe Dutang and Petr Savicky. *randtoolbox: Generating and Testing Random Numbers*. R package version 1.17. 2015.
- [Ext+16] Peter Exterkate, Patrick JF Groenen, Christiaan Heij, and Dick van Dijk. „Non-linear forecasting with many predictors using kernel ridge regression“. In: *International Journal of Forecasting* 32.3 (2016), pp. 736–753 (cit. on pp. 7, 15).
- [HK08] Robin Hyndman and Yeasmin Khandakar. „Automatic time series forecasting: The forecast Package for R“. In: *Journal of Statistical Software* 27.3 (2008), pp. 1–22.

- [HK70] Arthur E Hoerl and Robert W Kennard. „Ridge regression: Biased estimation for nonorthogonal problems“. In: *Technometrics* 12.1 (1970), pp. 55–67 (cit. on p. 7).
- [HS97] Sepp Hochreiter and Jürgen Schmidhuber. „Long short-term memory“. In: *Neural computation* 9.8 (1997), pp. 1735–1780.
- [JK08] S Joe and F Kuo. *Notes on generating Sobol sequences*. <http://web.maths.unsw.edu.au/~fkuo/sobol/joe-kuo-notes.pdf>. 2008.
- [Jür00] Manuela Jürgens. *LaTeX: eine Einführung und ein bisschen mehr*. FernUniversität Gesamthochschule in Hagen, 2000 (cit. on p. 3).
- [Jür95] Manuela Jürgens. *LaTeX: Fortgeschrittene Anwendungen*. FernUniversität Gesamthochschule in Hagen, 1995 (cit. on p. 3).
- [KJUM11] Markus Kohm and Jens-Uwe-Morawski. *KOMA-Script: Die Anleitung*. 2011 (cit. on p. 3).
- [Lüt05] Helmut Lütkepohl. *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005 (cit. on p. 14).
- [Mie11a] André Miede. *A Classic Thesis Style: An Homage to The Elements of Typographic Style*. 2011 (cit. on p. 3).
- [Nie92] Harald Niederreiter. *Random number generation and quasi-Monte Carlo methods*. SIAM, 1992.
- [NS87] Charles R Nelson and Andrew F Siegel. „Parsimonious modeling of yield curves“. In: *Journal of business* (1987), pp. 473–489 (cit. on p. 7).
- [Pan12] Alan Pankratz. *Forecasting with dynamic regression models*. Vol. 935. John Wiley & Sons, 2012.
- [Pao+94] Yoh-Han Pao, Gwang-Hoon Park, and Dejan J Sobajic. „Learning and generalization characteristics of the random vector functional-link net“. In: *Neurocomputing* 6.2 (1994), pp. 163–180.
- [Pen55] Roger Penrose. „A generalized inverse for matrices“. In: *Mathematical proceedings of the Cambridge philosophical society*. Vol. 51. 03. Cambridge Univ Press. 1955, pp. 406–413.
- [Pfa+08] Bernhard Pfaff et al. „VAR, SVAR and SVEC models: Implementation within R package vars“. In: *Journal of Statistical Software* 27.4 (2008), pp. 1–32.
- [Ren+16] Ye Ren, PN Suganthan, N Srikanth, and Gehan Amaratunga. „Random vector functional link network for short-term electricity load demand forecasting“. In: *Information Sciences* 367 (2016), pp. 1078–1093.
- [Rum+88] DE Rumelhart, GE Hinton, and RJ Williams. „Learning internal representations by error propagation“. In: *Neurocomputing: foundations of research*. MIT Press. 1988, pp. 673–695.
- [Sch+92] Wouter F Schmidt, Martin A Kraaijveld, and Robert PW Duin. „Feedforward neural networks with random weights“. In: *Pattern Recognition, 1992. Vol. II. Conference B: Pattern Recognition Methodology and Systems, Proceedings., 11th IAPR International Conference on*. IEEE. 1992, pp. 1–4.
- [Wic16] Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer, 2016.

- [ZS16] Le Zhang and PN Suganthan. „A comprehensive evaluation of random vector functional link networks“. In: *Information Sciences* 367 (2016), pp. 1094–1105.
- [App10a] Apple Inc. *Keynote '09 User Guide*. Apple Inc., 2010 (cit. on p. 3).
- [App10b] Apple Inc. *Numbers '09 User Guide*. Apple Inc., 2010 (cit. on p. 3).
- [App10c] Apple Inc. *Pages '09 User Guide*. Apple Inc., 2010 (cit. on p. 3).

Webseiten

- [@Mie11b] André Miede. *A Classic Thesis Style by André Miede*. 2011. URL: <http://www.miede.de/index.php?page=classicthesis> (visited on May 27, 2011) (cit. on p. 3).
- [@Fre10] Free Software Foundation, Inc. *GNU General Public License*. 2010. URL: <http://www.gnu.org/licenses/gpl.html> (visited on May 27, 2011) (cit. on p. 3).

List of Figures

2.1	Observed discount rates from Deutsche Bundesbank website, from 2002 to the end 2015	17
-----	--	----

List of Tables

2.1	Summary of observed discount rates from Deutsche Bundesbank website, from 2002 to the end 2015	16
-----	--	----

Colophon

This thesis was typeset with \LaTeX 2_ε. It uses the *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the *Clean Thesis* style at <http://cleanthesis.der-ric.de/>.

Declaration

You can put your declaration here, to declare that you have completed your work solely and only with the help of the references you mentioned.

City, August 26, 2015

Ricardo Langner