FUNDAÇÃO GETÚLIO VARGAS
ESCOLA DE ECONOMIA SÃO PAULO
MESTRADO PROFISSIONAL EM ECONOMIA E FINANÇAS

APRENDIZADO DE MÁQUINA PARA ANÁLISE DE CURVAS DE JUROS

Luis Cláudio De Giovanni Pache de Faria

Orientador: Prof. Marcelo Fernandes

SEMINÁRIOS DE DISSERTAÇÃO II

OBJETIVO

Executar rotinas do modelo Kernel-Ridge não paramétrico de aprendizado de máquina para projeção de curvas de juros, proposta por Filipovitch, Pelger e Ye (2022), e compará-la aos modelos paramétricos correntemente usados, mais notavelmente o de Nelson, Siegel e Svensson (1994) dentro do contexto de títulos públicos brasileiros.

PERGUNTA

Como o modelo Kernel-Ridge se compara a modelos paramétricos para projeção de curva de juros de títulos públicos brasileiros sob o ponto de vista de usabilidade, precisão e robustez?

CONTRIBUIÇÃO

Aplicar localmente um modelo moderno de aprendizado de máquina para projeção de curva de juros dentro do ambiente de títulos públicos brasileiros e divulgar esta nova metodologia no nosso meio acadêmico.

REVISÃO DA LITERATURA – DE POOTER (2007)

Definições preliminares

- Taxa de juros futura instantânea
- Taxa de juros a termo
- Curva de desconto
- Relação biunívoca

$$\lim_{\tau^* \downarrow 0} f_t(\tau, \tau^*) = f_t(\tau)$$

$$y_t(\tau) = \frac{1}{\tau} \int_0^{\tau} f_t(m) \mathrm{d}m$$

$$P_t(\tau) = \exp\left[-\tau y_t(\tau)\right]$$

$$f_t(\tau) = -\frac{1}{P_t(\tau)} \frac{\mathrm{d}P_t(\tau)}{\mathrm{d}\tau} = y_t(\tau) + \tau \frac{\mathrm{d}y_t(\tau)}{\mathrm{d}\tau}$$

Modelos Paramétricos

- Nelson e Siegel (1987)

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t} \left[\frac{1 - \exp\left(-\frac{\tau}{\lambda_t}\right)}{\left(\frac{\tau}{\lambda_t}\right)} \right] + \beta_{3,t} \left[\frac{1 - \exp\left(-\frac{\tau}{\lambda_t}\right)}{\left(\frac{\tau}{\lambda_t}\right)} - \exp\left(-\frac{\tau}{\lambda_t}\right) \right]$$

- Svensson (1994)

$$y_t(\tau) = \beta_{1,t} + \beta_{2,t} \left[\frac{1\text{-}\exp\left(-\frac{\tau}{\lambda_{1,t}}\right)}{\left(\frac{\tau}{\lambda_{1,t}}\right)} \right] + \beta_{3,t} \left[\frac{1\text{-}\exp\left(-\frac{\tau}{\lambda_{1,t}}\right)}{\left(\frac{\tau}{\lambda_{1,t}}\right)} - \exp\left(-\frac{\tau}{\lambda_{1,t}}\right) \right] + \beta_{4,t} \left[\frac{1\text{-}\exp\left(-\frac{\tau}{\lambda_{2,t}}\right)}{\left(\frac{\tau}{\lambda_{2,t}}\right)} - \exp\left(-\frac{\tau}{\lambda_{2,t}}\right) \right]$$

Modelos não Paramétricos

- Liu e Wu (2021)

$$\hat{g}_{LW}(x) = \frac{\sum_{n=1}^{360} K_{h(x)}(n-x) \exp\left(-(y_n + (x-n) y'_n) x\right)}{\sum_{n=1}^{360} K_{h(x)}(n-x)}$$

$$K_{h(x)}(n-x) = \frac{1}{\sqrt{2\pi h(x)^2}} \exp\left(-\frac{(n-x)^2}{2h(x)^2}\right)$$

- Filipovic, Pelger e Ye (2022): apresentado adiante

REVISÃO DA LITERATURA – THEODORIDIS (2015)

Regressão Ridge - Penalizar função de perda quadrática

$$\hat{\theta}_{\mathrm{LS}} = \arg\min_{\theta} \sum_{n=1}^{N} (y_n - \theta^{\mathrm{T}} x_n)^2.$$

$$\begin{aligned} \text{minimize:} \quad & J(\theta) = \sum_{n=1}^{N} \left(y_n - \theta^T x_n\right)^2, \\ \text{subject to:} \quad & \|\theta\|^2 \le \rho, \end{aligned}$$

$$\text{minimize:} \quad L(\theta, \lambda) = \sum_{n=1}^{N} \left(y_n - \theta^T x_n\right)^2 + \lambda \|\theta\|^2 : \quad \text{Ridge Regression.}$$

$$\hat{\theta}_R = \arg\min_{\theta} \left\{ \|y - X\theta\|^2 + \lambda \|\theta\|^2 \right\}$$

Reproducing kernel Hilbert Spaces (RKHS)

Consideremos um espaço linear $\mathbb{H}$ de funções de valores reais definidas em um conjunto $\chi \subseteq \mathbb{R}^l$ tal que $\mathbb{H}$ seja um espaço Hilbert, ou seja, que exista um operador de produto interno $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ que defina a norma correspondente $\| \cdot \|_{\mathbb{H}}$, e que $\mathbb{H}$ seja completo em relação a esta norma. Um espaço Hilbert $\mathbb{H}$ é chamado de *reproducing kernel Hilbert space* (RKHS) se existe uma função $\kappa : \chi \times \chi \to \mathbb{R}$ tal que para cada $x \in \chi$, $\kappa(\cdot, x)$ pertence a $\mathbb{H}$, e além disso $\kappa(\cdot, \cdot)$ possui a chamada *reproducing property*:

$$f(\boldsymbol{x}) = \langle f, \kappa(\cdot, \boldsymbol{x}) \rangle, \ \forall f \in \mathbb{H}, \ \forall \boldsymbol{x} \in \mathcal{X} : \quad \text{Reproducing Property.}$$

$$\mathcal{X} \ni \boldsymbol{x} \longmapsto \phi(\boldsymbol{x}) := \kappa(\cdot, \boldsymbol{x}) \in \mathbb{H} : \quad \text{Feature Map,}$$

$$\langle \phi(\boldsymbol{x}), \phi(\boldsymbol{y}) \rangle = \kappa(\boldsymbol{x}, \boldsymbol{y}) : \quad \text{Kernel Trick.}$$

Regressão Kernel Ridge

Regressão não linear: $y_n = g(x_n) + \eta_n, \ n = 1, 2, \ldots, N.$

Teorema da representação: f estimador de g em RKHS $f(\boldsymbol{x}) = \sum_{n=1}^{N} \theta_n \kappa(\boldsymbol{x}, \boldsymbol{x}_n).$

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}),$$

$$J(\boldsymbol{\theta}) := \sum_{n=1}^{N} \left(y_n - \sum_{m=1}^{N} \theta_m \kappa(\boldsymbol{x}_n, \boldsymbol{x}_m) \right)^2 + C\langle f, f \rangle,$$

$$\hat{y} = \sum_{n=1}^{N} \hat{\theta}_n \kappa(\boldsymbol{x}, \boldsymbol{x}_n) = \hat{\boldsymbol{\theta}}^{\mathrm{T}} \boldsymbol{\kappa}(\boldsymbol{x}),$$

$$\hat{y}(\boldsymbol{x}) = \boldsymbol{y}^{\mathrm{T}} (\mathcal{K} + CI)^{-1} \boldsymbol{\kappa}(\boldsymbol{x}).$$

REVISÃO DA LITERATURA – FILIPOVIC, PELGER, YE (2022)

Problema Fundamental

Preços de títulos, sendo C = fluxos de caixa com dimensão $M \times N$ (M datas e N títulos)

$$P_i^g = \sum_{j=1}^{N} C_{ij} g(x_j).$$

$$P_i = P_i^g + \epsilon_i.$$

Otimização

$$\min_g \left\{ \sum_{i=1}^{M} \omega_i \left(P_i - P_i^g\right)^2 \right\}$$

restrição (suavidade = sem arbitragem)

$$\|g\|_{\alpha,\delta} = \left(\int_0^\infty \left(\delta g'(x)^2 + (1-\delta) g''(x)^2\right) \mathrm{e}^{\alpha x}\, dx \right)^{\frac{1}{2}}$$

$$\min_{g \in \mathcal{G}_{\alpha,\delta}} \left\{ \underbrace{\sum_{i=1}^{M} \omega_i (P_i - P_i^g)^2}_{\text{pricing error}} + \lambda \underbrace{\|g\|_{\alpha,\delta}^2}_{\text{smoothness}} \right\}$$

com pesos

$$\omega_i = \frac{1}{M} \frac{1}{(D_i P_i)^2}$$

Solução Geral

$$\hat{g}(x) = 1 + \sum_{j=1}^{N} k(x, x_j) \beta_j,$$

com $\beta = C^\top (CKC^\top + \Lambda)^{-1} (P - C\mathbf{1}),$ matriz *kernel* $\mathcal{K}_{ij} = k(x_i, x_j)$ $\mathrm{N} \times \mathrm{N}$, $\Lambda = \mathrm{diag}(\lambda/\omega_1, \ldots, \lambda/\omega_M)$

$\alpha = 0, \delta \in (0, 1)$:

$$k(x,y) = \frac{1}{\delta}(x \wedge y) + \frac{1}{2\delta\rho}\left(\mathrm{e}^{-\rho(x+y)} - \mathrm{e}^{\rho(x\wedge y) - \rho(x \vee y)}\right)$$

$$\rho = \sqrt{\delta/(1-\delta)}$$

$\alpha = 0, \delta = 1$:

$$k(x,y) = x \wedge y$$

$\alpha > 0, \delta = 1$:

$$k(x,y) = -\frac{\alpha}{\delta \ell_2^2}\left(1 - \mathrm{e}^{-\ell_2 x} - \mathrm{e}^{-\ell_2 y}\right) + \frac{1}{\alpha\delta}\left(1 - \mathrm{e}^{-\alpha(x \wedge y)}\right) + \frac{1}{\delta\sqrt{D}}\left(\frac{\ell_1^2}{\ell_2^2}\mathrm{e}^{-\ell_2(x+y)} - \mathrm{e}^{-\ell_1(x\wedge y) - \ell_2(x \vee y)}\right)$$

$$D = \alpha^2 + 4\delta/(1-\delta),\ \ell_1 = \tfrac{\alpha - \sqrt{D}}{2},\ and\ \ell_2 = \tfrac{\alpha + \sqrt{D}}{2}$$

$\alpha > 0, \delta = 1$:

$$k(x,y) = \frac{1}{\alpha}\left(1 - \mathrm{e}^{-\alpha(x \wedge y)}\right)$$

$(\alpha, \delta) = (0, 0)$ não é especificado

(*) Notação $a \wedge b = \min\{a,b\}$ / $a \vee b = \max\{a,b\}$

BASE DE DADOS

Histórico de Preços e Taxas dos Títulos Públicos Federais

- https://www.tesourodireto.com.br/titulos/historico-de-precos-e-taxas.htm
- http://www.tesourotransparente.gov.br/ckan/dataset/taxas-dos-titulos-ofertados-pelo-tesouro-direto
- Tesouro prefixado LTN – 2002 a 2022 (21 arquivos)
- Tesouro prefixado com juros semestrais NFN-F – 2004 a 2022 (19 arquivos)
- Dados usados: Data base, Tipo de título, Vencimento, Preço PU

Data Frame

- 42.292 registros
- 4.681 datas base (janeiro de 2004 a novembro de 2022) em 213 meses
- 59 títulos (tipo e vencimento)
- *Sampling* – Datas base no último dia útil de cada mês (213 datas)

Filtros de outliers sugeridos pela literatura - Gürkaynak, Sack e Wright (2007) e Filipovic, Pelger e Ye (2022)

- Títulos com vencimento inferior a 90 dias
- Curva NSS - Erros acima de 3 desvios padrão das taxas baseado na seção cruzada temporal
- Curva KR - Erros acima de 3 desvios padrão das taxas baseado na seção cruzada temporal
- Estimação dos modelos com e sem filtros (Modelo KR robusto a outliers)

Simulação (resampling) – Filipovic, Pelger e Ye (2022)

- Suposição de taxas no próximo dia útil t+1 (teste) são virtualmente iguais àquelas em t (treinamento)

METODOLOGIA

1. Compilação dos dados históricos de preços de títulos soberanos brasileiros pré-fixados;

2. Construção de rotinas em R e Python para tratamento dos dados de preços históricos;

3. Execução de otimização da função paramétricas Nelson-Siegel-Svensson em R através do método por algoritmo genético (aka *Differential Evolution*). Considerado também o método do Gradiente (aka *Steepest Descent*);

4. Execução de aprendizado de máquina com rotinas em Python para o modelo Kernel-Ridge com os hiper-parâmetros λ, α e δ fixos;

5. Tratamento e análise de resultados com rotinas em R considerando erros médios quadráticos (RMSE) ponderados pela duração, além de medidas de tensão e suavidade dos modelos NSS e KR, através de discretização numérica:

   RMSE $\mathrm{RMSE} = \frac{1}{T}\sum_{t=1}^{T} \mathrm{RMSE}_t, \quad \mathrm{RMSE}_t = \sqrt{\sum_{i=1}^{M_t} \omega_{i,t} \left(P_{i,t} - \hat{P}^g_{i,t}\right)^2}.$ com $\omega_i = \frac{1}{M}\frac{1}{(D_i P_i)^2}$

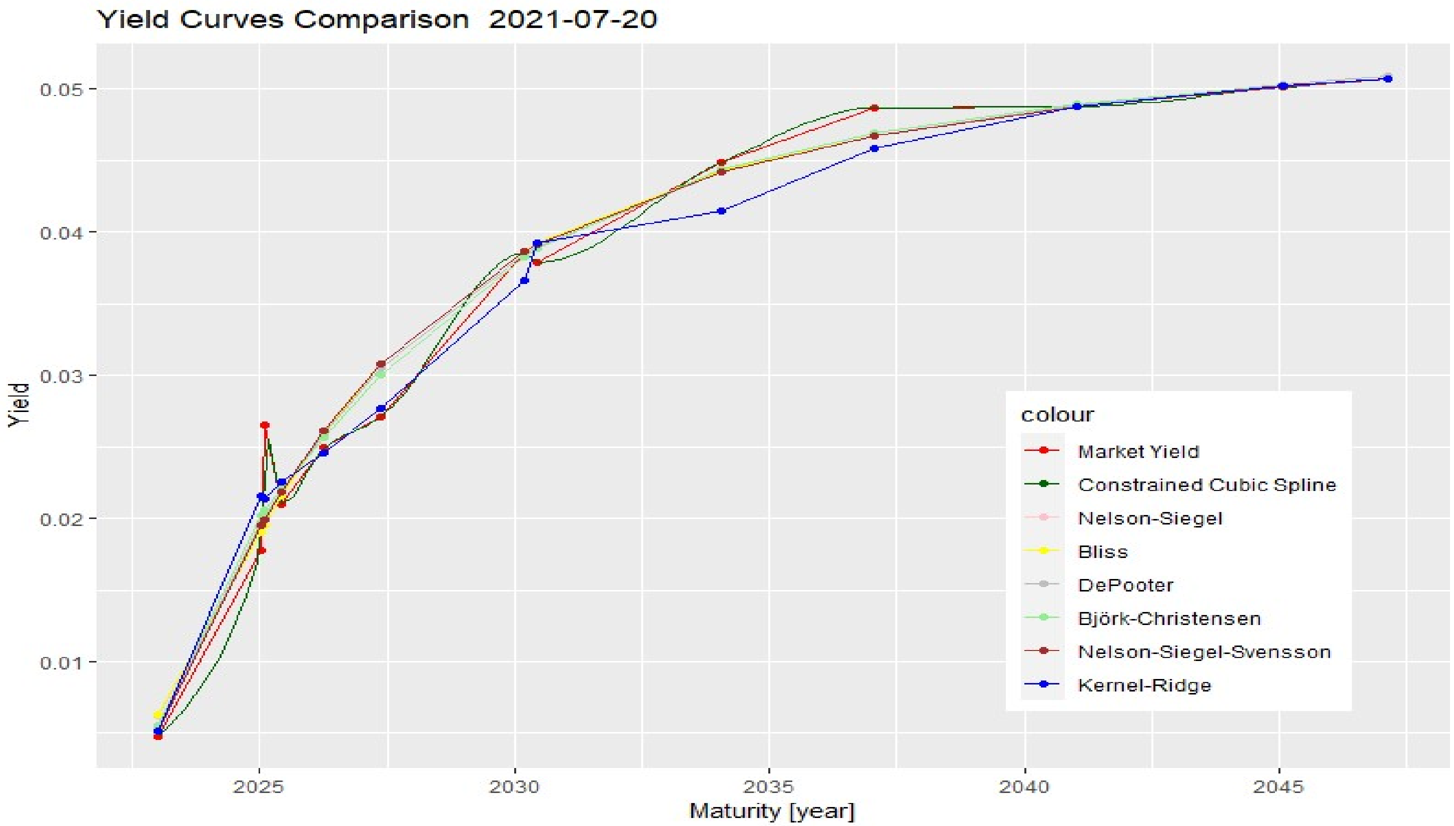
   Tensão e Suavidade: $\frac{1}{|\mathcal{R}|}\int_{\mathcal{R}} g'(x)^2\, dx$ e $\frac{1}{|\mathcal{R}|}\int_{\mathcal{R}} g''(x)^2\, dx$

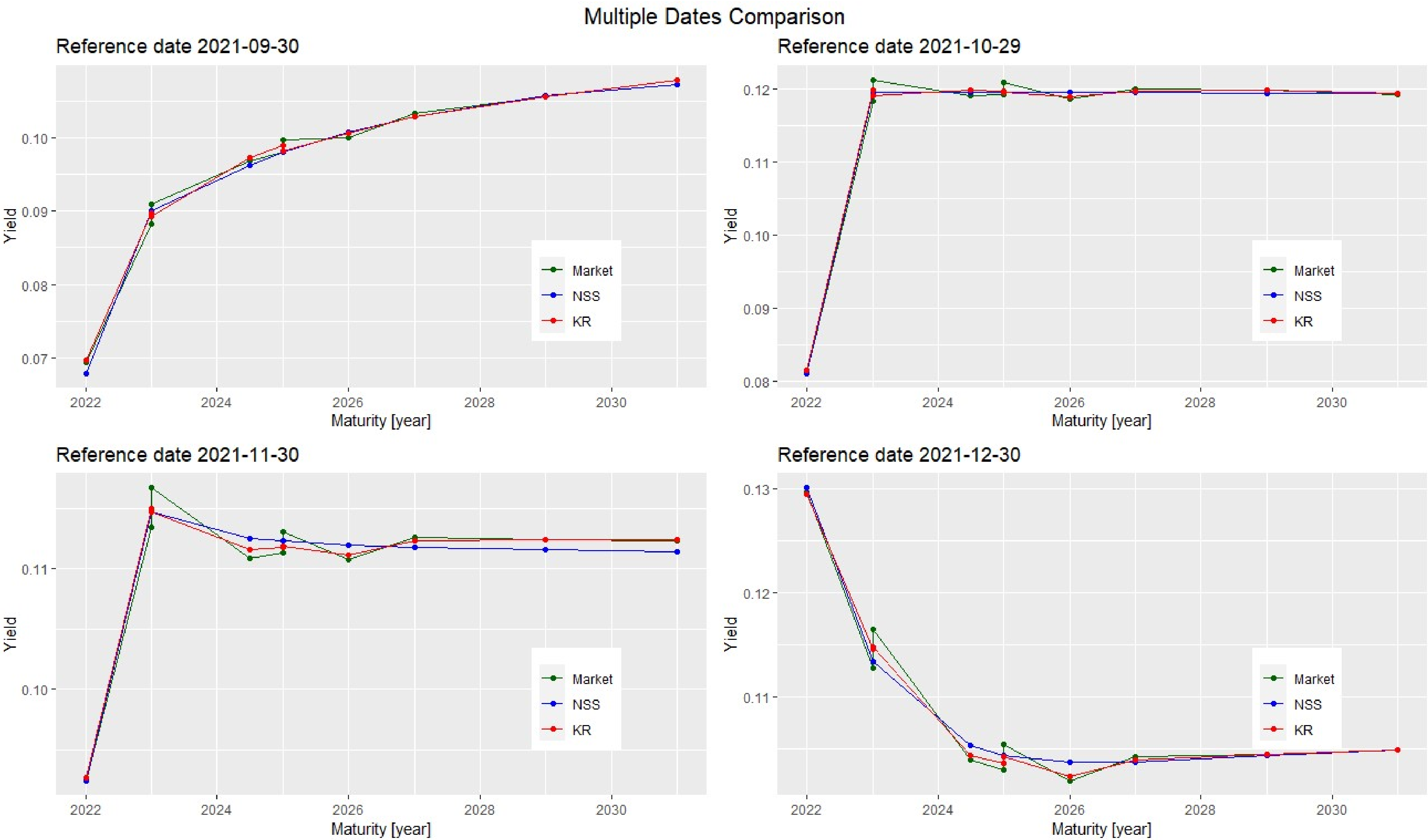
6. Confronto de cada modelo (NSS vs KR) com apresentação de conclusões, virtudes e limitações do modelo KR à luz de nossa realidade.

RESULTADOS PRELIMINARES

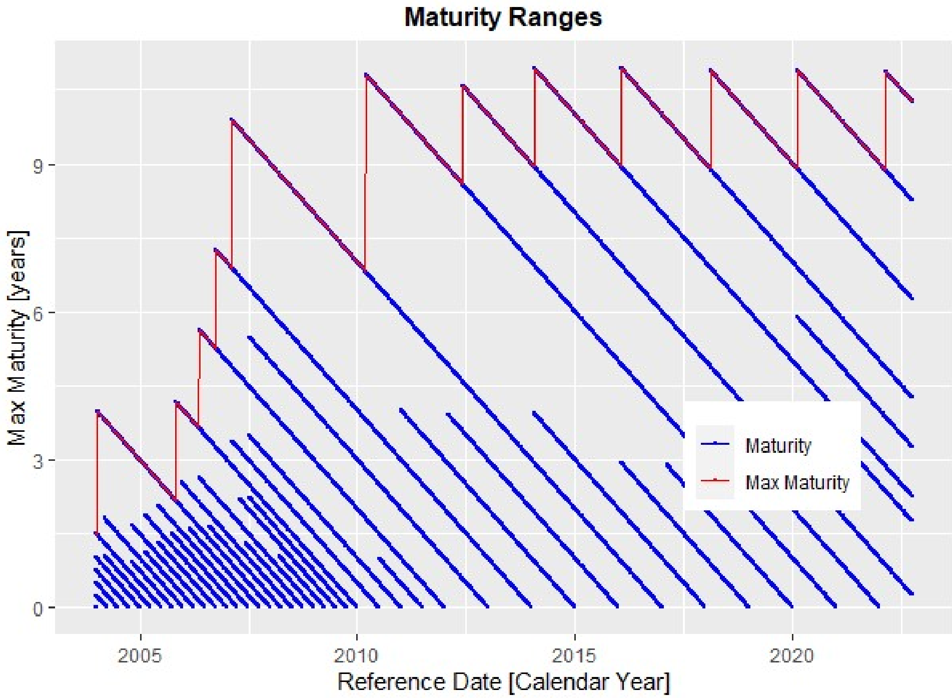
MODELOS ANALISADOS

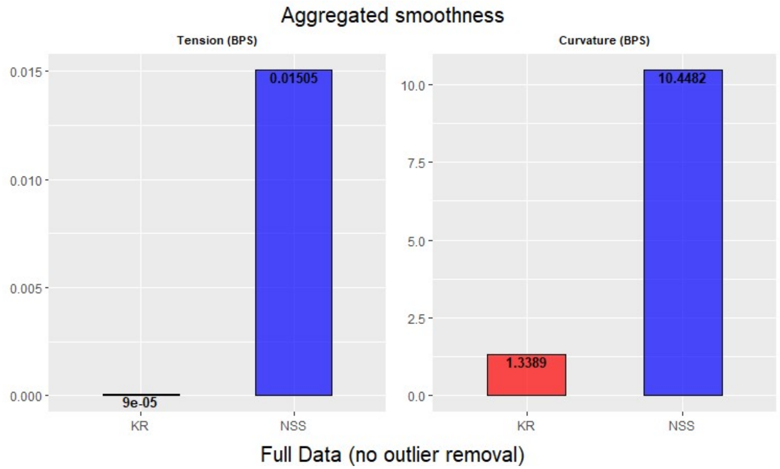
- **Interpolação por Constrained Cubic Spline**
- **Paramétricos: Nelson-Siegel, Bliss, DePooter, Bjork-Christensen, Nelson-Siegel-Svensson (otimização por Algoritmos Genéticos)**
- **Não Paramétricos: Filipovic-Pelger-Ye (otimização por aprendizado de máquina Kernel-Ridge)**

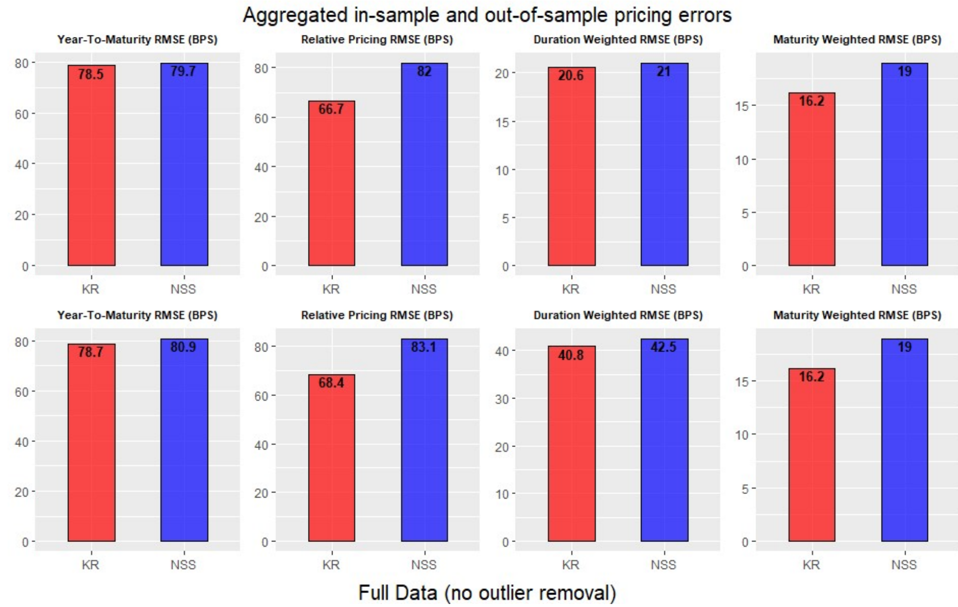


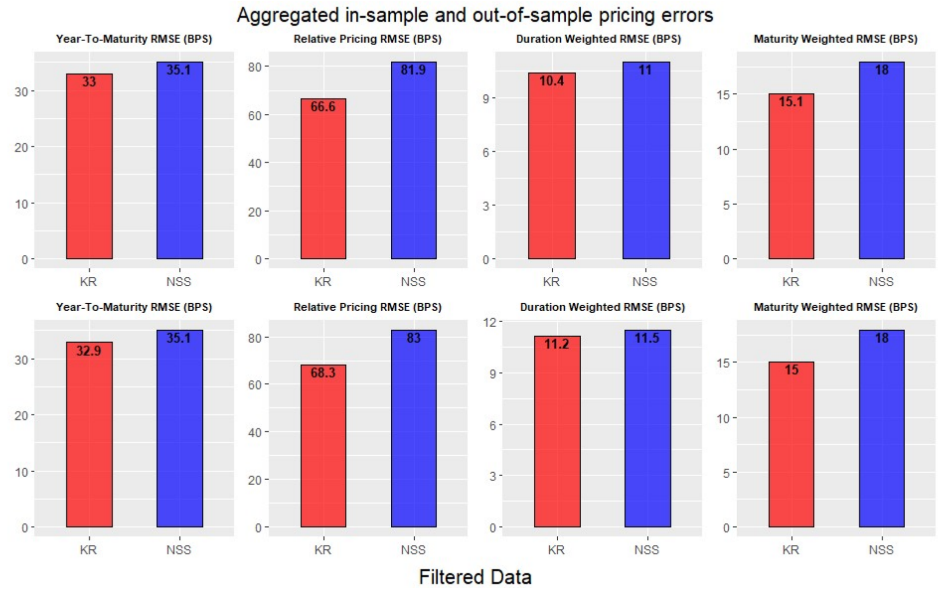


Full Data (no outlier removal)

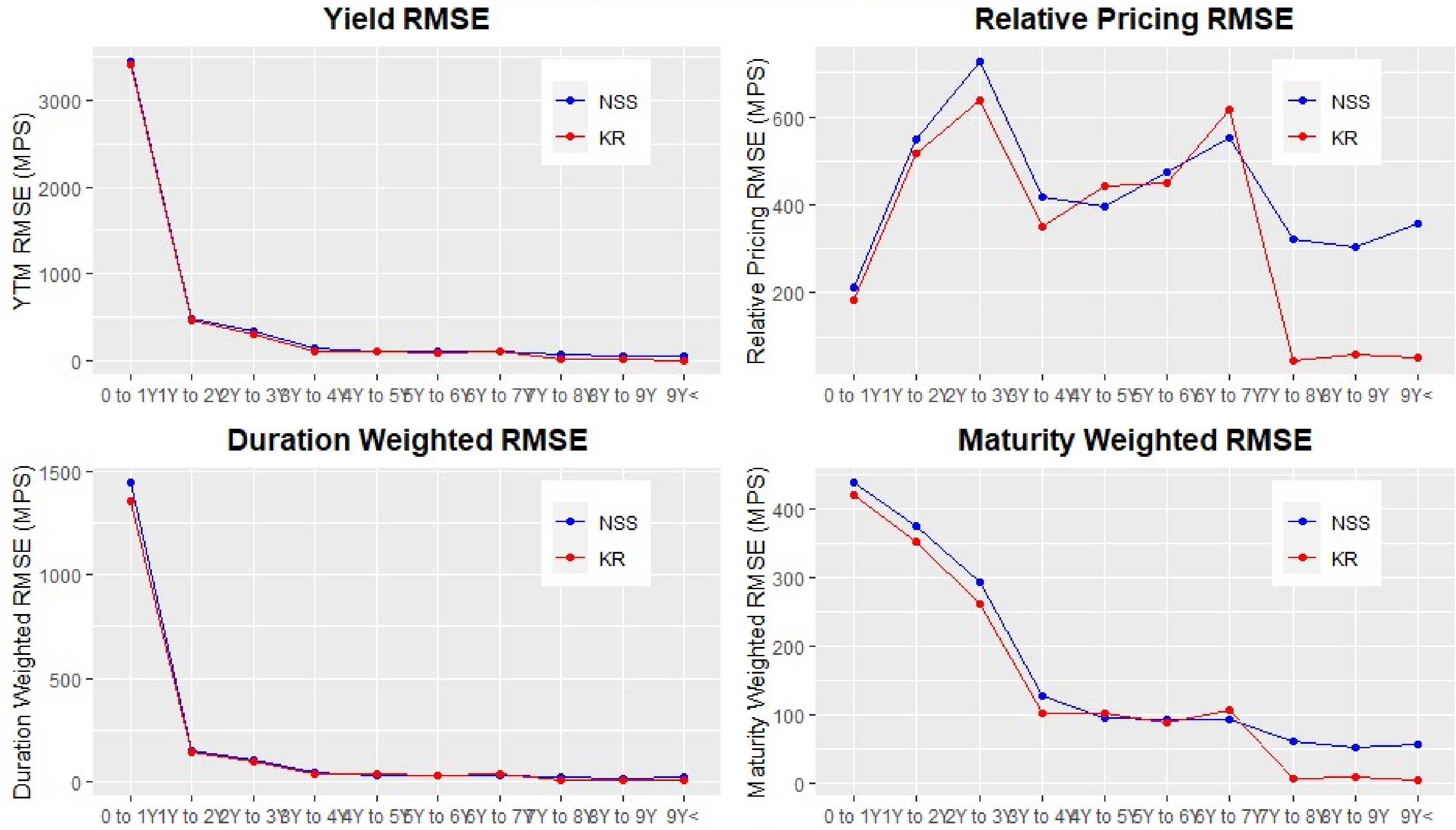

Maturity Ranges
Max Maturity [years]
Reference Date [Calendar Year]
2005
2010
2015
2020
Maturity
Max Maturity


Aggregated smoothness
Tension (BPS)
Curvature (BPS)
9e-05
0.01505
1.3389
10.4482
KR
NSS
Full Data (no outlier removal)


Aggregated in-sample and out-of-sample pricing errors
Year-To-Maturity RMSE (BPS)
Relative Pricing RMSE (BPS)
Duration Weighted RMSE (BPS)
Maturity Weighted RMSE (BPS)
78.5
79.7
66.7
82
20.6
21
16.2
19
78.7
80.9
68.4
83.1
40.8
42.5
16.2
19
KR
NSS
Full Data (no outlier removal)


Aggregated in-sample and out-of-sample pricing errors
Year-To-Maturity RMSE (BPS)
Relative Pricing RMSE (BPS)
Duration Weighted RMSE (BPS)
Maturity Weighted RMSE (BPS)
33
35.1
66.6
81.9
10.4
11
15.1
18
32.9
35.1
68.3
83
11.2
11.5
15
18
KR
NSS
Filtered Data

Out-of-sample pricing errors for different maturities



Full Data (no outlier removal)