



Universidade de Brasília

Instituto de Ciências Exatas

Departamento de Estatística

Dissertação de Mestrado

**Uma modelagem estatística do tempo até a  
aposentadoria dos servidores estatutários civis do  
poder executivo da União**

por

**Igor Ribeiro Mendonça**

**Orientador: Prof. Dr. Guilherme Souza Rodrigues**

Brasília, 03 de dezembro de 2021

# **Uma modelagem estatística do tempo até a aposentadoria dos servidores estatutários civis do poder executivo da União**

**por**

**Igor Ribeiro Mendonça**

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade de Brasília, como parte dos requisitos necessários para obtenção do título de Mestre em Estatística.

Comissão julgadora:

Prof. Dr. Guilherme Souza Rodrigues (Orientador) - EST/UnB

Dr. Claudio Hamilton Matos dos Santos (Membro externo) - IPEA

Prof. Dr. Raul Yukihiro Matsushita (Membro Interno) - EST/UnB

Profa. Dra. Thais Carvalho Valadares Rodrigues (Suplente) - EST/UnB

Brasília, 03 de dezembro de 2021

## RESUMO

Este trabalho se propõe a criar um modelo linear para estimar o tempo até a aposentadoria dos servidores estatutários civis da União. A partir de uma base de dados cadastrais fornecida e gerada em agosto de 2019 pela Secretária de Previdência do Governo e utilizando o algoritmo *Generalized Additive Model for Location Scale and Shape* (GAMLSS). Os dados cadastrais foram tratados para filtrar os servidores ativos aptos a aposentar em cada ano de 2005 a 2018. Procedimento que gera censuras nos tempos até a aposentadoria observados, o que foi tratado na modelagem proposta. O modelo foi então comparado quantitativamente com outros tipos de arquitetura, como redes neurais, e qualitativamente com modelos da literatura que propuseram a modelar a aposentadoria também. Por fim, o valor estimado até o tempo até a aposentadoria é tratado para levar em conta regra previdenciária da idade compulsória, que o servidor é obrigado a se aposentar.

**Palavras-chave:** Modelos lineares, aposentadoria, redes neurais, dados censurados, análise de sobrevivência.

## ABSTRACT

This work suggests a linear model to estimate the time until retirement of civilian public workers of Brazil's Government. The data used was provided and generated in August of 2019 by the Department of Retirement of Brazil and the algorithm used in the work is *Generalized Additive Model for Location Scale and Shape* (GAMLSS). The registered data were treated to filter the workers that were active and had acquired the right to retire in each year from 2005 until 2018. The procedure to this filter produces censorship in the time until retirement observed in the data, which was treated in the development of the model. Then, the final model was compared quantitatively with other kinds of algorithms, such as neural nets and deep learning models, and compared qualitatively with other proposals from the literature that tried to model retirement as well. At last, the time estimated to retirement is treated to accommodate a retirement law that requires the public worker to retire when a certain age is reached.

**Keywords:** Linear models, retirement, neural nets, deep learning, censored data, survival analysis.

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>6</b>
1.1	Trabalhos correlatos . . . . .	8
1.2	Proposta e desafios . . . . .	11
<b>2</b>	<b>Contexto técnico e teórico</b>	<b>13</b>
2.1	<i>Generalized Additive Models for Location Scale and Shape</i> (GAMLSS) . . . . .	13
<b>3</b>	<b>Materiais</b>	<b>14</b>
<b>4</b>	<b>Metodologia de análise</b>	<b>16</b>
4.1	Ajustes realizados nas bases de dados para aplicar modelos . . . . .	17
4.2	Modelagem GAMLSS . . . . .	19
4.3	Redes neurais . . . . .	21
<b>5</b>	<b>Resultados</b>	<b>25</b>
5.1	Modelos GAMLSS . . . . .	25
5.2	Modelos alternativos . . . . .	30
5.3	Modelos da literatura . . . . .	32
5.4	Suavização dos coeficientes estimados pelos modelos GAMLSS . . . . .	35
5.5	Truncamento dos valores estimados pelo modelo GAMLSS . . . . .	38
<b>6</b>	<b>Considerações finais</b>	<b>40</b>
	<b>Referências</b>	<b>42</b>
	<b>Apêndice</b>	<b>45</b>

# 1 Introdução

O modelo atual de previdência no Brasil é o de repartição: quem contribui custeia os benefícios daqueles servidores em inatividade. É financiado pelos trabalhadores com carteira assinada que contribuem para um fundo, pelo empregador e, por último, pelo governo Federal, que também participa garantindo que não ocorra um desequilíbrio financeiro no sistema.

Esse sistema previdenciário se divide em três: Regime Geral de Previdência Social (RGPS), Regime Próprio de Previdência Social (RPPS) e a Previdência Complementar. Os servidores do Governo Federal contribuem para o RPPS da União. Estes servidores estão distribuídos nos três Poderes, executivo (civis e militares), legislativo e judiciário, e no ministério público.

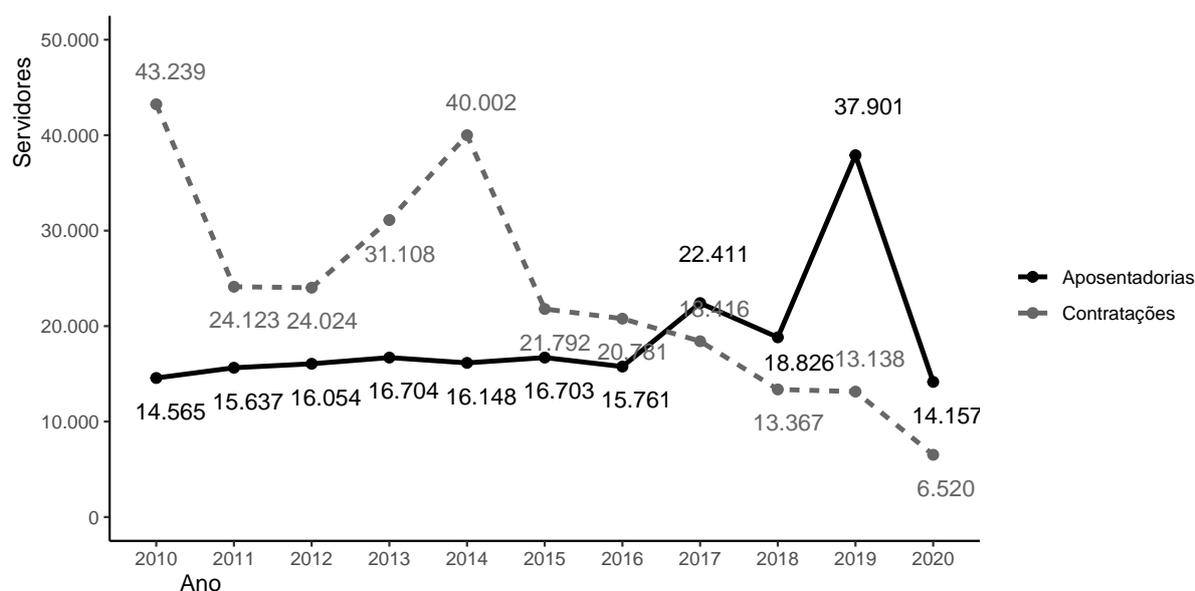


Figura 1 - Quantidade de contratações de servidores estatutários no serviço público e servidores estatutários aposentados no serviço público da União de 2010 a 2019. Dados extraídos do Painel Estatístico de Pessoal do Ministério da Economia.

Estima-se que existam por volta de 556 mil servidores estatutários trabalhando no poder executivo da União (excluindo-se judiciário, ministério público e legislativo) e 428 mil servidores

do executivo já aposentados pela União<sup>1</sup>. Em relação aos gastos, os servidores estatutários do executivo civil (excluindo militares) representam quase 70% do total gasto com pessoal em 2020 (185 bilhões de 266 bilhões de Reais)<sup>2</sup>. Isso importa porque, pela constituição federal, a equipe econômica do Governo Federal só tem autonomia nos gastos com esse pessoal. Os demais poderes (e Ministério Público) têm autonomia com seus gastos, e os militares apresentam planos de carreira próprios.

A situação corrente dos servidores estatutários da União é um sistema em que a cada 100 servidores ativos existem por volta de 77 servidores inativos. Portanto, é imprescindível saber quando novos inativos entrarão no sistema previdenciário, porque isso representa menos contribuintes e mais pessoas se beneficiando dos fundos previdenciários. Tendo em vista ainda que, atualmente, o ritmo de contratação de novos servidores está caindo, que representariam novos contribuintes. Foram contratados aproximadamente 42 mil servidores estatutários em 2010 e apenas 6 mil em 2019 (Figura 1), ao passo que o ritmo de aposentadorias vem crescendo nos últimos anos, em especial em 2019 (Figura 1).

A pesquisa de Schettini e Pires (2018) dá conta que aproximadamente 25% dos servidores estatutários da União já estavam aptos a se aposentar em 2016 (137 mil de 543 mil servidores). Observa-se que na prática, os servidores não se aposentam logo que adquirem o direito a aposentadoria. De acordo com os dados do Ministério da Economia, aproximadamente 16 mil servidores se aposentaram naquele ano, ou seja, apenas 11% dos servidores aptos a se aposentar o fizeram. De certa forma, pela lei, há um incentivo financeiro para permanência no cargo assim que a elegibilidade é adquirida. O servidor adquire o direito ao abono permanência, o que significa o fim da contribuição previdenciária.

---

<sup>1</sup>Dados da avaliação atuarial do Projeto de Lei de Diretrizes Orçamentárias (PLDO) de 2021 relativos a situação em 31 de dezembro de 2019. Disponível em <https://www.gov.br/economia/pt-br/assuntos/orcamento/orcamento/orcamentos-anuais/2021/pldo/projeto-de-lei-de-diretrizes-orcamentarias>

<sup>2</sup>Dados extraídos do Painel Estatístico de pessoal do Ministério da Economia. Disponível em <https://www.gov.br/economia/pt-br/acesso-a-informacao/servidores/servidores-publicos/painel-estatistico-de-pessoal>

## 1.1 Trabalhos correlatos

Schettini, Santos e Pires (2018) abordam, de certa maneira, o desafio de estimar quando servidores irão se aposentar. O artigo se concentra em microssimulações não comportamentais para estimar o impacto dos gastos com aposentadorias e pensões dos servidores públicos civis do Executivo Federal resultantes da PEC 287, que começou a tramitar no congresso em dezembro de 2016.

Para tanto, o trabalho se concentrou em calcular quando os servidores (as) ficariam elegíveis. Este cálculo é feito com base na idade, data de entrada no serviço público, tempo de contribuição atual e anterior ao serviço público, gênero e cargo de cada servidor e com base nas regras previdenciárias do artigo 40 da Constituição Federal, da Emenda Complementar nº 41 artigo 2º e artigo 6º e da Emenda Complementar nº 47 artigo 3º.

Os autores estimaram que muitos servidores continuam trabalhando após estarem aptos a aposentadoria usufruindo do abono permanência, chamados riscos iminentes ou expirados. O modelo proposto estima a probabilidade de aposentadoria de acordo com a idade de cada servidor a partir de uma distribuição beta.

O texto de projeções de gasto com pessoal de 2020 a 2060 de Santos, Pires e Mendonça (2020) aborda a aposentadoria no serviço público estadual (civil e militar) como objeto de modelagem. As regras de aposentadorias pré-reforma e pós-reforma da previdência aprovada no Congresso Nacional, Proposta de Emenda à Constituição 6/2019, são utilizadas.

O trabalho também utilizou algumas informações como idade, data de entrada no serviço público, tempo de contribuição total, gênero e cargo dos servidores para estimar com base nas regras previdenciárias as respectivas datas de elegibilidade. Porém, é utilizada a técnica estatística de Análise de Sobrevivência para modelar a probabilidade de aposentadoria.

Casalecchi, Orair e Souza (2020) também fazem projeções de gasto com pessoal para os servidores civis da União de 2020 a 2060 sob as diferentes regras previdenciárias (anteriores a PEC 6/2019 e considerando a reforma aprovada). A transição para a inatividade não supõe

aleatoriedade. O ano provável de aposentadoria é o ano em que o servidor atinge a idade compulsória (75 anos de idade) ou quando o mesmo cumpre os critérios de elegibilidade para a aposentadoria voluntária acrescido de um tempo médio estimado de uso do abono permanência. O que acontecer primeiro. Além disso, o modelo leva em conta que o ativo pode se aposentar por invalidez com uma probabilidade de 0,16%, o que não foi tratada em Santos, Pires e Mendonça (2020). É importante ressaltar que o texto baseia sua metodologia de ano provável de aposentadoria com base em relatórios produzidos pelo próprio governo, o “Relatório de Avaliação atuarial do RPPS da União”<sup>3</sup> e o “Relatório de estudo Atuarial dos impactos da PEC nº 6/2019 – RPPS dos Servidores Cíveis da União”.<sup>4</sup>

O trabalho de Bhering (2005) aplica simulações estocásticas de acumulação de recursos nos planos de previdência para projetar a nível individual as rendas de suas aposentadoria. Diferentemente do proposto nesse trabalho, a aleatoriedade do momento da aposentadoria não é considerada. Os indivíduos aposentam com idades pré-fixadas e benefício vitalício resultante do montante capitalizado durante o período de contribuição previdenciária.

A literatura internacional é vasta no estudo da transição para a aposentadoria. Especialmente na área de economia do trabalho e atuarial. Vale citar Kenc (2000) que desenvolveu um modelo para a incerteza na decisão de aposentadoria sob a ótica de teorias econômicas, modelo intertemporal de consumo, oferta de emprego e na escolha do portfólio do fundo de pensão. O autor visa derivar a idade ótima de aposentadoria com base em fatores de trabalho, remuneração e o seu benefício previdenciário.

Destacam-se dois artigos que empregam técnicas estatísticas e de aprendizagem de máquinas, Salazar e Penas (2019) e Samwick (1998). O primeiro artigo utiliza dados de

---

<sup>3</sup>Hipóteses metodológicas referidas desenvolvidas por grupo de trabalho composta por técnicos e especialistas do próprio governo (SPREV, STN, SOF, SEPLAN, e SGP). Disponível em <https://www.gov.br/previdencia/pt-br/assuntos/previdencia-no-servico-publico/demonstrativos/grupo-de-trabalho-resultado-financeiro-e-atuarial-do-rpps-da-uniao>

<sup>4</sup>Disponível em [http://sa.previdencia.gov.br/site/2019/04/1-RPPS-DA-UNIAO\\_RELATORIO-DO-ESTUDO-ATUARIAL-DOS-IMPACTOS-DA-PEC-No-6-2019.pdf](http://sa.previdencia.gov.br/site/2019/04/1-RPPS-DA-UNIAO_RELATORIO-DO-ESTUDO-ATUARIAL-DOS-IMPACTOS-DA-PEC-No-6-2019.pdf)

previdência privada do México para prever se os indivíduos se aposentarão ou não em função de características pessoais dos indivíduos e variáveis macroeconômicas. São testados modelos de classificação. O segundo artigo, Samwick (1998), utiliza dados de uma pesquisa sobre o histórico de aposentadoria americana (RHS – *Retirement Historic Survey*). O estudo tenta prever se um indivíduo optou por se aposentar ou não usando um modelo estatístico. Mais detalhes sobre estes dois textos serão aprofundados na Seção 5.3.

O artigo de Stock e Wise (1990) se propôs a modelar o custo de oportunidade de se aposentar em detrimento de acumular mais riqueza no futuro - isto é, mais contribuições para o plano de previdência. A estimativa se dá por meio de análise numérica com regras de aposentadoria.

No campo de projeções de populações, também na literatura internacional, Dekkers *et al.* (2015) usa modelos de transição de eventos sociais (aposentadoria, educação, casamento, invalidez, trabalho) para avaliar o impacto do envelhecimento da classe trabalhadora na estrutura previdenciária da Bélgica, Suécia e Hungria a partir de bases administrativas. E Fredrikson (1998) aplica um modelo de microsimulação dinâmica de corte transversal usando bases públicas da Noruega para entender a composição populacional, as consequências do nível de escolaridade, disponibilidade de mão de obra e os benefícios de aposentadoria.

Tang e Burr (2014) utilizaram dados públicos americanos com foco na passagem da classe trabalhadora para a aposentadoria de homens e mulheres com idades entre 51 e 74 anos. Os autores ajustaram uma regressão multinomial logito com dados socio demográficos, saúde, família e força de trabalho para avaliar como cada fator impacta na verossimilhança do indivíduo estar em diferentes grupos – aposentado integral, aposentado parcialmente ou trabalhando em meio expediente, trabalhador em tempo integral, trabalhador portador de deficiência física ou dona(o) de casa.

## 1.2 Proposta e desafios

A fim de melhor compreender como as aposentadorias acontecem no Brasil, essa dissertação se propõe a tentar explicar a nível individual (servidor) quais fatores influenciam as aposentadorias. Por exemplo, iremos quantificar os efeitos do gênero do servidor, sua idade, tempo de serviço, região do país entre outros fatores no tempo até a aposentadoria. Embora não seja o foco, o modelo possibilitará prever, para cada servidor, a distribuição de probabilidade do tempo até a aposentadoria de acordo com suas características. Modelagem que difere do estimado em Salazar e Penas (2019) e Samwick (1998). Apesar destes modelos utilizarem variáveis importantes, o desenho das análises se limitam a prever se o indivíduo aposenta ou não. O nível de detalhe do servidor é notadamente maior que o assumido em Schettini, Santos e Pires (2018); Santos, Pires e Mendonça (2020) e; Casalecchi, Orair e Souza (2020). Isso proporciona maior variabilidade no tempo estimado até a aposentadoria por considerar diversos fatores individuais do servidor e conseqüentemente, obtemos uma melhor explicação dos fatores decisivos para aposentadoria. Ou seja, uma melhor compreensão do ritmo de novas aposentadorias.

Precisamente, vamos modelar o tempo que os servidores estatutários civis da União tomam para se aposentar a partir da data de elegibilidade a aposentadoria em função de um conjunto de variáveis sócio demográficas, como escolaridade, região, gênero, idade e outras. Os dados utilizados são formados a partir de uma base bruta administrativa pública fornecida pela Secretária de Previdência. Em seguida, queremos investigar o comportamento temporal dessas relações aplicando um meta modelo sobre os parâmetros estimados do primeiro modelo minimizando as variações de estimativa dos parâmetros de cada modelo em função dos anos.

A conjuntura norte americana de aposentadoria é, naturalmente, diferente em relação ao Brasil – suas regras previdenciárias e estruturas de aposentadoria, por exemplo. Porém há inúmeros artigos que estudam a influência de fatores do indivíduo (ou mesmo macroeconômicos) na tomada de decisão da aposentadoria como gênero, idade, raça, saúde, estado civil, conjuntura econômica, educação, tempo de serviço entre outros exemplos. O que sustenta a importância

de relacioná-las ao tempo até aposentadoria. Ver Pienta, Burr e Mutchler (1994); Wong e Hardy (2009); Wright (2012); Warner, Hayward e Hardy (2010); Choudhury e Bridges (2009), Mutchler *et al.* (1999), Kim e DeVaney (2005); Maestas (2010); Mutchler *et al.* (1997); Brown e Warner (2008) e; Rogowski e Karoly (2000).

Há alguns desafios para estimação que este trabalho tenta solucionar como: 1) a decisão individual de cada servidor em quando se aposentar pode superar o alcance da base gerando censuras (servidores que adquiriram o direito a aposentadoria mas ainda não o fizeram até o momento de geração da base bruta utilizada); 2) os servidores são obrigados a se aposentarem ao atingir 75 anos de idade, chamada aposentadoria compulsória, gerando truncamento para ser modelado na variável resposta; 3) não há informação confiável disponível acerca de quantos anos os servidores trabalharam antes de entrarem no serviço público. Informação de grande relevância para o cálculo de elegibilidade de aposentadoria por meio das regras previdenciárias. Muitos servidores só averbam seus trabalhos anteriores na véspera do momento da aposentadoria. De acordo com Schettini, Santos e Pires (2018), apenas aproximadamente 25% dos servidores ativos averbaram seus trabalhos anteriores e; 4) não há dados cadastrais de boa qualidade.

O texto é dividido em mais cinco Seções. A seguir, na Seção 2, apresentamos uma revisão das principais técnicas estatísticas e computacionais utilizadas neste projeto. A base de dados adotada é apresentada em detalhes na Seção 3. A metodologia é descrita na Seção 4, incluindo os filtros, limpezas e ajustes na base de dados. Por fim, os resultados são apresentados na Seção 5 e as considerações finais na Seção 6. Além disso, as referências e apêndices foram disponibilizados ao final do texto. Os códigos desenvolvidos estão disponíveis sob consulta.

## 2 Contexto técnico e teórico

### 2.1 *Generalized Additive Models for Location Scale and Shape (GAMLSS)*

GAMLSS são modelos estatísticos de regressão de distribuição univariada apresentados por Rigby e Stasinopoulos (2001) e Rigby e Stasinopoulos (2005). A implementação permite modelar até quatro parâmetros a depender da família de distribuição assumida,  $\mu, \sigma, \nu$  e  $\tau$ , a partir das variáveis explicativas. Esse conjunto de modelos nasceu como forma de flexibilizar a suposição de distribuição exponencial dos modelos GLM (*Generalized Linear Models*).

Nos modelos GAMLSS, a suposição de família exponencial para a variável resposta é relaxada e substituída por uma família de distribuição mais abrangente, o que inclui famílias assimétricas e/ou achatadas. Estimativa de máxima verossimilhança é utilizada para ajustar os modelos.

O modelo assume que as observações  $y_i$  ( $i = 1, 2, \dots, n$ ) são independentes e seguem uma função densidade de probabilidade  $f(y_i|\theta_i)$  condicional em  $\theta_i$ , em que  $\theta_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ip})$  é um vetor de  $p$  parâmetros, que são associados as variáveis explicativas. Assume-se que os parâmetros são definidos por  $\mu_i, \sigma_i, \nu_i, \tau_i$ , em que  $\mu_i$  e  $\sigma_i$  são ditos parâmetros de locação e escala, respectivamente, enquanto  $\nu_i$  e  $\tau_i$  são parâmetros de forma. Sendo  $y$ , vetor de tamanho  $n$ , variável resposta e  $g_k(\cdot)$  função monotônica de ligação, em relação ao parâmetro  $\theta_k$ , com as variáveis explicativas. O modelo é dito aditivo semi paramétrico e dado por:

$$\begin{aligned}g_1(\mu) &= \eta_1 = \mathbf{X}_1\beta_1 + \sum_{i=1}^{J_1} h_{j1}(x_{j1}) \\g_2(\sigma) &= \eta_2 = \mathbf{X}_2\beta_2 + \sum_{i=1}^{J_2} h_{j2}(x_{j2}) \\g_3(\nu) &= \eta_3 = \mathbf{X}_3\beta_3 + \sum_{i=1}^{J_3} h_{j3}(x_{j3}) \\g_4(\tau) &= \eta_4 = \mathbf{X}_4\beta_4 + \sum_{i=1}^{J_4} h_{j4}(x_{j4}),\end{aligned}$$

Em que  $\mu, \sigma, \nu, \tau, \eta_k$  e  $x_{jk}$ , para  $j = 1, 2, \dots, J_k$  e  $k = 1, 2, 3, 4$ , são vetores de tamanho  $n$ .  $h_{jk}$  é uma função não paramétrica aditiva das variáveis explicativas  $x_{jk}$  (pode incluir termos de suavização, *splines*, termos de coeficientes variáveis, também pode ser generalizado para incluir efeitos aleatórios entre outros) avaliadas em  $x_{jk}$ , que são assumidos fixos e conhecidos assim como a matriz  $\mathbf{X}_k$ . Enquanto  $\beta_k$  são os vetores dos parâmetros. Se, para  $k = 1, 2, 3, 4$ ,  $J_k = 0$ , então o modelo reduz para um modelo completamente paramétrico dado por:

$$g_k(\theta_k) = \eta_k = \mathbf{X}_k \beta_k$$

O componente paramétrico  $\mathbf{X}_k \beta_k$  pode incluir termos lineares e interações para as variáveis explicativas e também fatores, polinômios e polinômios fracionados.

O modelo é ajustado maximizando a função de verossimilhança dada por  $l$ :

$$l = \sum_{i=1}^n \log f(y_i | \theta_i)$$

A censura na variável resposta também é implementada no GAMLSS. Supondo uma observação com censura à direita de uma variável resposta, sua contribuição para a função de log verossimilhança  $l$  é:  $\log [1 - F(y|\theta)]$ . Em que  $F(y|\theta)$  é a função de distribuição acumulada da variável resposta observada naquele ponto censurado.

A dissertação foi feita utilizando o software estatístico R<sup>5</sup> utilizando as bibliotecas *gamlss* e toda a sua estrutura para a modelagem, ver Rigby e Stasinopoulos (2007). Os pacotes *data.table* Dowle e Srinivasan (2021), *tidyverse* Wickham *et al.* (2019) e *lubridate* Grolemund e Wickham (2011) foram utilizados para a manipulação das bases e ajustes.

### 3 Materiais

A base de dados usada neste trabalho foi gerada – em julho de 2019 – pela Secretária de Previdência do governo em um trabalho desenvolvido por um grupo técnico sobre Regime Próprio

---

<sup>5</sup><https://www.r-project.org/>

de Previdência Social (RPPS) formado por pesquisadores e interessados do setor previdenciário, como servidores do Instituto de Pesquisa Econômica Aplicada (IPEA) e da própria Secretária. A base é identificada e composta por registros pessoais (nome, data de nascimento, escolaridade, estado civil) de todos os servidores ativos, aposentados, dependentes e pensionistas que tiveram vínculo empregatício com a União em algum momento. Também inclui registros funcionais expondo município do exercício, data de entrada no serviço público, saída, inatividade, data de abono, bem como registros financeiros relacionados a remuneração, contribuição ou valor do benefício. Vale ressaltar a variável “DT\_OCOR\_ABONO\_PERM\_PSS” que informa a data estimada de aquisição do direito ao abono permanência (Tabela 1). É, de fato, crucial para este trabalho porque é considerada a data em que o servidor se tornou elegível para a aposentadoria.

A Tabela 1 exemplifica algumas das principais variáveis utilizadas, suas definições e tipos. Ao total, a base apresenta 305 variáveis e 2.960.754 linhas. Cada linha pode ser entendida como um vínculo empregatício.

Variável	Tipo	Observação
SG_REGIME_JURIDICO	Caractere	Sigla do regime jurídico do vínculo do servidor
DT_OCOR_INGR_SPUB_SERV	Data	Data da ocorrência de ingresso do servidor no serviço público federal
DT_OCOR_INATIVIDADE_SERV	Data	Data da ocorrência da aposentadoria
DT_OBITO	Data	Data do óbito do servidor falecido
DT_OCOR_ABONO_PERM_PSS	Data	Data de aquisição do direito ao abono permanência
DT_NASC_SERV	Data	Data de nascimento do servidor
LIQUIDO	Númerico	Valor da remuneração líquida do servidor
NO_ORGAO	Caractere	Nome do órgão do servidor
NO_NATUREZA_JURIDICA	Caractere	Nome da natureza jurídica do órgão do servidor
UF_UORG	Caractere	UF da unidade de exercício do servidor
NO_ESTADO_CIVIL	Caractere	Nome do estado civil do servidor
NO_NIVEL_ESCOLARIDADE	Caractere	Nome do nível de escolaridade do servidor
NO_DEFICIENCIA_FISICA	Caractere	Nome da deficiência física do servidor
CO_JOR_TRAB_CARGO	Númerico	Código da jornada de trabalho do cargo do servidor

Tabela 1 - Dicionário das principais variáveis da base utilizada no estudo.

A base é fundamental para o trabalho dado seu nível profundo de detalhe (Tabela 1) e tam-

bém por permitir identificar os servidores que estavam ativos e aposentados em cada ano com a metodologia descrita na Seção 4. Com isso, foi criado um painel com informações detalhadas do quadro de servidores ativos e inativos da União em cada ano, o que é inédito<sup>6</sup>. Além disso, é possível identificar os servidores que vieram a óbito ou mesmo aqueles que entraram no serviço público. O resultado dos estoques e fluxos ano a ano pode ser observado na Tabela 2. Importante ressaltar que não é necessário restringir ao intervalo temporal de 2005 a 2018 descrito na Tabela 2.

O referido filtro para obter os servidores estatutários ativos em um dado ano  $x$  é feito com base em 7 variáveis e pode ser traduzido como: 1) servidores do regime jurídico estatutário e; 2) que não aposentaram antes de 1º de janeiro de  $x$  e; 3) não saíram do serviço público antes de 1º de janeiro de  $x$  e; 4) que não faleceram antes de 1º de janeiro de  $x$  e, por último; 5) entraram no serviço público antes de 1º de janeiro de  $x$ .

A lógica para obter os servidores estatutários ingressantes na União em um ano  $x$  se dá por obter servidores do regime jurídico estatutário na base e que tenham sido admitidos em  $x$ . Os Falecidos seguem a mesma lógica. Enquanto os aposentados são aqueles servidores estatutários que aposentaram em  $x$  mas não faleceram em  $x$ .

## 4 Metodologia de análise

A proposta do trabalho é modelar o tempo que o servidor leva para se aposentar a partir do momento em que se torna elegível a aposentadoria. Por isso, optou-se por filtrar os servidores ativos em 1º de janeiro de cada ano de análise, que se tornaram aptos a se aposentar naquele dado ano. Além dos filtros para reconhecer os servidores ativos descritos na Seção anterior, foi utilizado um filtro adicional para reconhecer a elegibilidade do servidor.

---

<sup>6</sup>Santos, Brito e Cavalcante (2020) tenta replicar essa mesma estrutura de dados em painel para os servidores públicos estaduais. Foi criado um painel com informações dos servidores ativos estaduais de 2014 a 2018 a partir da RAIS. Disponível em: [https://rshinypainel.shinyapps.io/shiny\\_test/](https://rshinypainel.shinyapps.io/shiny_test/)

Ano	Ativos	Ingressantes	Falecidos	Aposentados	Fluxo Líquido
2005	483.176	17.419	1.333	6.253	9.833
2006	488.125	15.134	1.398	7.351	6.385
2007	505.061	29.602	1.348	9.334	18.920
2008	504.779	13.799	1.458	10.866	1.475
2009	511.141	22.269	1.456	14.184	6.629
2010	525.991	34.674	1.441	14.796	18.437
2011	543.958	40.486	1.428	15.694	23.364
2012	544.694	23.217	1.437	16.000	5.780
2013	544.672	22.969	1.364	16.630	4.975
2014	549.752	28.958	1.373	16.212	11.373
2015	566.975	42.079	1.371	16.812	23.896
2016	563.131	20.096	1.348	15.854	2.894
2017	561.091	19.677	1.429	22.435	-4.187
2018	550.204	17.124	1.338	18.825	-3.039

Tabela 2 - Quantidade de servidores estatutários, aposentados, ingressantes e falecidos por ano na União de 2005 a 2018 de acordo com a base bruta.

Para efeito de comparação com os dados atuariais públicos, de acordo com a avaliação atuarial do PLDO de 2019, válido como quantitativo de dezembro de 2017, são 554 mil servidores estatutários. Valor próximo do observado na base em 2018. O quantitativo referente a dezembro de 2016 (comparável com a base de 2017) de acordo com o PLDO de 2018 é equivalente a 578 mil.

#### **4.1 Ajustes realizados nas bases de dados para aplicar modelos**

A variável “DT\_OCOR\_ABONO\_PERM\_PSS” da base é utilizada como o ano de elegibilidade. Como dito anteriormente, a variável informa a data que o servidor adquiriu o direito ao abono permanência. Filtra-se, portanto, aqueles servidores que se tornaram aptos a aposentadoria no ano x. Os servidores que apresentavam datas faltantes nesta variável tiveram o seu

Ano	Ativos	Ativos aptos a se aposentar	Percentual de aptos
2005	483.176	8.509	1.76%
2006	488.125	11.087	2.27%
2007	505.061	13.617	2.7%
2008	504.779	18.691	3.7%
2009	511.141	21.775	4.26%
2010	525.991	21.723	4.13%
2011	543.958	20.963	3.85%
2012	544.694	20.593	3.78%
2013	544.672	19.067	3.5%
2014	549.752	21.034	3.83%
2015	566.975	20.162	3.56%
2016	563.131	21.974	3.9%
2017	561.091	20.321	3.62%
2018	550.204	18.067	3.28%

Tabela 3 - Quantidade de servidores estatutários e aptos a aposentadoria por ano de 2005 a 2018.

provável ano de elegibilidade estimado levando em conta as regras previdenciárias do artigo 40 da Constituição Federal, da Emenda Complementar nº 41 artigo 2º e artigo 6º e da Emenda Complementar nº 47 artigo 3º e as informações de idade, tempo de serviço, gênero e cargo do servidor. Foi criada uma covariável explicativa para identificar os servidores aptos a se aposentar que foram filtrados pela data de abono estimada. O total de servidores ativos e servidores aptos a se aposentar por ano encontrados na base são dispostos na Tabela 3.

Foi decidido por restringir a análise do estudo entre 2005 e 2018 (coluna “Ativos aptos a se aposentar” das Tabelas 3 e 4). O ano de 2005 é o ano da última reforma previdenciária, Emenda Complementar nº 47 de 2005, enquanto o ano de 2018 é o último ano completo antes da geração da base bruta. A Tabela 4 mostra a distribuição dos casos censurados em cada ano – servidores que não se aposentaram até o momento de consolidação da base bruta em julho de 2019. Vale notar que 23% (1981 servidores) dos ativos que se tornaram aptos em 2005 não exerceram o direito a aposentadoria por ao menos 14 anos. Por outro lado, a grande maioria dos servidores

que se tornaram aptos a aposentadoria em 2018 (80%) optou por não se aposentar até julho de 2019.

Ano	Ativos aptos a se aposentar	Casos censurados	Proporção de censurados
2005	8.509	1.981	23.3%
2006	11.087	2.839	25.6%
2007	13.617	3.866	28.4%
2008	18.691	5.386	28.8%
2009	21.775	6.951	31.9%
2010	21.723	7.764	35.7%
2011	20.963	8.196	39.1%
2012	20.593	8.828	42.9%
2013	19.067	9.101	47.7%
2014	21.034	10.973	52.2%
2015	20.162	11.335	56.2%
2016	21.974	14.030	63.8%
2017	20.321	14.340	70.6%
2018	18.067	14.452	80.0%

Tabela 4 - Quantidade de servidores estatutários aptos a aposentadoria por ano de 2005 a 2018.

## 4.2 Modelagem GAMLSS

Após os processamentos na base, ajustamos modelos GAMLSS em cada um dos anos separadamente. A variável resposta é o tempo até a aposentadoria (em meses) calculada da data que o servidor adquiriu o direito a se aposentar até a data limite da criação da base, julho de 2019. A variável resposta é tratada como em análise de sobrevivência com uma variável indicadora de censura (a direita). Aqueles servidores que se aposentaram até julho de 2019 tiveram seus tempos até a aposentadoria observados e aqueles que optaram por não se aposentar até a data limite são considerados casos censurados (distribuição das censuras na Tabela 4). As variáveis explicativas disponíveis para modelagem são: gênero do servidor, tempo de serviço público, idade, variável indicadora se o servidor é professor, variável indicadora se a data de abono foi

estimada, remuneração líquida, natureza jurídica (administração direta federal, autarquia especial ou federal ou fundação federal), região do país que o servidor exerce o trabalho, estado civil – solteiro (a), viúvo (a), divorciado (a) e casado (a) –, escolaridade (fundamental, médio, superior, mestrado ou doutorado), jornada de trabalho em horas, variável indicadora de pessoa com deficiência, variável indicadora se o servidor entrou por concurso e, por fim, se o servidor se encontra na última classe padrão do seu cargo/carreira.

Cinco conjuntos de modelos foram ajustados considerando a família da distribuição binomial negativa. Cada conjunto se refere a um tipo de modelo GAMLSS aplicado (Tabela 5) (variando o conjunto de variáveis e parâmetros estimados, no caso  $\mu$  e  $\sigma$ , média e dispersão) nas bases filtradas por ano (2005 a 2018). Foram ajustados cinco tipos de modelos nesses 14 anos de 2005 a 2018, o que resultam em 70 ajustes diferentes.

Modelo	Média	Dispersão
Nulo	Somente intercepto	Somente intercepto
Trivial apenas média	Todas variáveis	Somente intercepto
Trivial - Média e variância	Todas variáveis	Todas variáveis
Trivial com stepwise para média e variância	Todas variáveis	Todas variáveis
Stepwise com interação para média e variância	Todas variáveis e suas interações	Todas variáveis e suas interações

Tabela 5 - Resumo dos modelos GAMLSS e as variáveis utilizadas em cada parâmetro.

O primeiro conjunto de modelos GAMLSS, um pra cada ano de 2005 até 2018, é ajustado sem covariáveis, apenas o intercepto para os parâmetros de média  $\mu$  e de escala  $\sigma$  (modelo inicial). O segundo conjunto considera as variáveis listadas acima para ajustar apenas o parâmetro de média  $\mu$  (modelo trivial). A média e o desvio padrão foram modeladas considerando as mesmas variáveis listadas acima para o terceiro conjunto de modelos. Para o quarto conjunto, foi aplicado um método de seleção de variáveis<sup>7</sup> para obter o melhor modelo, com base na me-

<sup>7</sup>O procedimento se chama *Stepwise selection*, o que consiste em selecionar o conjunto de variáveis que melhor ajusta aos dados por meio de uma medida de informação, no caso AIC. A seleção é feita começando com o modelo nulo, sem variáveis explicativas. O próximo passo é aplicar o modelo com cada uma das variáveis disponibilizadas e verificar qual apresenta o melhor resultado em relação ao AIC (menor valor). A variável que apresentar o melhor

dida AIC<sup>8</sup>, em cada ano. Foi disponibilizada todas as variáveis, sem interações ou efeitos não lineares, para que o melhor modelo fosse ajustado. Nesse caso, o ajuste foi feito para a média  $\mu$  e para o parâmetro de escala  $\sigma$ . Por último, também foi aplicado um método de seleção de variáveis para obter o melhor modelo, com base na medida AIC, ajustando os dados em cada ano considerando todas as variáveis e suas interações para a média  $\mu$  e o parâmetro de escala  $\sigma$ .

Os cinco tipos de modelos foram analisados em relação a medida de comparação BIC<sup>9</sup> e a parcimônia - quantidade de variáveis - a fim de escolher um modelo. Ou seja, aquele que explique bem os dados ao passo que tenha interpretabilidade.

### 4.3 Redes neurais

Propomos aplicar uma Rede Neural simples (sem camadas internas) para comparar com o modelo GAMLSS escolhido na etapa descrita anteriormente (as mesmas variáveis foram consideradas). O pacote *Keras*<sup>10</sup> Allaire e Chollet (2021) desenvolvido para o software R permite implementar funções de perda criadas pelo usuário para ajustar uma rede neural. Para se ter modelos comparáveis adaptamos a log-verossimilhança da distribuição binomial negativa tratando a censura, ver Nelson (1990), como a função de perda da rede neural. Antes de implementar a rede neural, validamos a função adaptada de perda com censura utilizando os recursos de otimização modelo é mantida e aplica-se o modelo adicionando cada uma das outras variáveis para verificar qual é o melhor modelo em relação ao AIC, inclusive o do passo anterior. O processo termina quando não há mais redução no valor do AIC. O modelo final não necessariamente inclui todas as variáveis.

<sup>8</sup>*Akaike Information Criterion* é um critério de informação utilizado na literatura estatística para avaliar o ajuste dos modelos. É usado também para comparação entre modelos. A fórmula para calcular seu valor é dada por  $AIC = 2k - 2l$  em que,  $k$  é o número de variáveis explicativas e  $l$  é a estimativa da log verossimilhança do modelo.

<sup>9</sup>*Bayesian Information Criterion* também é um critério de informação utilizado para seleção de modelos que é intimamente relacionado ao AIC. A sua fórmula se dá por  $AIC = \ln(n)k - 2l$  em que,  $k$  é o número de variáveis explicativas,  $\ln(n)$  é o logaritmo natural do número de observações  $n$  e  $l$  é a estimativa da log verossimilhança do modelo.

<sup>10</sup><https://keras.rstudio.com/>

do pacote *stats* R Core Team (2021) do R. Foi utilizado o otimizador L-BFGS-B, introduzido por Byrd *et al.* (1995). Chamamos o modelo de GAMLSS Manualmente. Esse procedimento foi desenvolvido para validar a função de perda e confirmar que obteríamos previsões similares ao modelo GAMLSS do pacote *gamlss* do R.

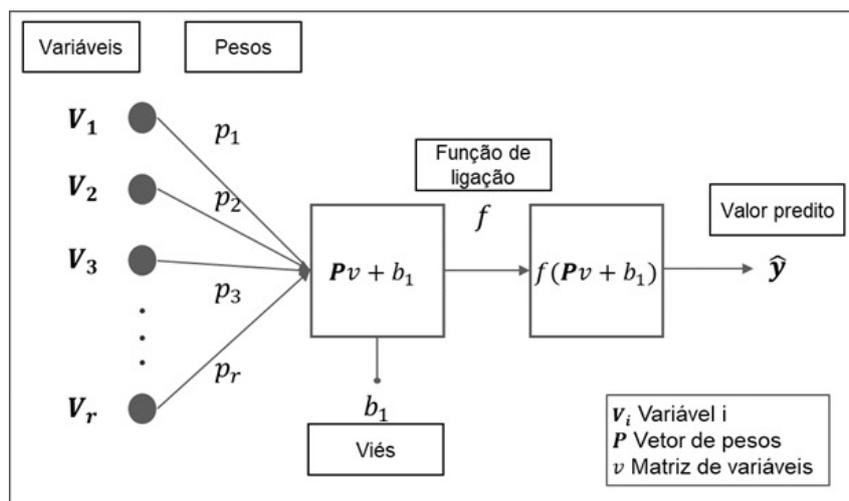


Figura 2 - Exemplo de arquitetura de rede neural simples (sem camadas internas)

Vale mencionar que em uma Rede Neural simples como na Figura 2, cada neurônio aplica um produto interno entre as variáveis e os pesos, adiciona o viés, aplica uma função de ligação e retorna o valor predito. O conjunto de neurônios forma uma camada do modelo. O modelo é treinado otimizando os pesos e vieses para minimizar uma função de perda utilizando um processo chamado *Gradient Descent*. Esse processo se dá basicamente nos seguintes passos: 1) pesos e viés iniciais são atribuídos, *forward propagation*; 2) calcula-se as previsões; 3) as previsões são utilizadas para calcular a função de perda para mensurar o erro da rede; 4) o algoritmo de *Gradient Descent* atualiza os pesos e vieses calculando os gradientes usando o método *backward propagation*. Esse processo é repetido até que os pesos e vieses não se alterem mais ou até que as diferenças entre esses valores fiquem abaixo de um valor previamente determinado.

Por isso, minimizando o negativo da log-verossimilhança no processo de otimização da

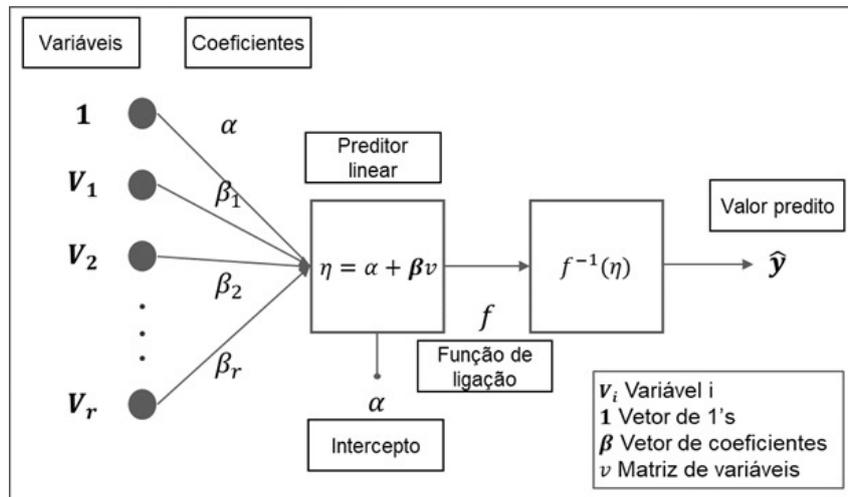


Figura 3 - Exemplo de arquitetura de modelo GAMLSS

rede neural equivale a maximizar a verossimilhança do modelo. Assim podemos utilizar as estimativas (valores preditos) para calcular o valor do log da verossimilhança dos dois modelos e comparar os dois métodos. Em particular, a rede neural otimiza os pesos e vieses ( $P$  e  $b_1$  da Figura 2) com o algoritmo *Root Mean Squared Propagation*<sup>11</sup> que é um tipo de algoritmo otimizador *Gradient Descent* mencionado. Ao passo que o GAMLSS otimiza os coeficientes do modelo ( $\beta$  e  $\alpha$  da Figura 3) utilizando o algoritmo RS, que é uma generalização do proposto por Rigby e Stasinopoulos (1996a,b), também equivale a maximizar a verossimilhança do modelo.

Além dos dois tipos de modelo utilizarem otimizadores diferentes, há questões relevantes a serem debatidas sobre como as variáveis são tratadas em cada modelo. O processo natural para melhorar os resultados de um modelo GAMLSS, como em Modelos Lineares Generalizados, inclui: testar diferentes variáveis, interações entre variáveis, termos quadráticos de variáveis, termos não lineares (GAMLSS permite essa inclusão) e modelar a dispersão, assim como a média. Estas são algumas alternativas. Métodos como o *Stepwise selection* são indicados nesses processos.

<sup>11</sup>Algoritmo não publicado. Proposto inicialmente no curso *Neural Networks for Machine Learning*, Coursera, por Geoffrey Hinton.

Essas estratégias combinadas permitem ganhos consideráveis em certos casos. Redes Neurais, no entanto, não necessitam a aplicação desses procedimentos. Devido a sua arquitetura, redes neurais mais complexas (variando quantidade de neurônios e camadas internas) podem capturar essas relações entre as variáveis. Por exemplo, um neurônio que liga uma variável para uma camada interna pode ser zerado (durante a otimização) retirando o efeito dessa variável na Rede neural. Isso equivale a não inclusão da variável no modelo.

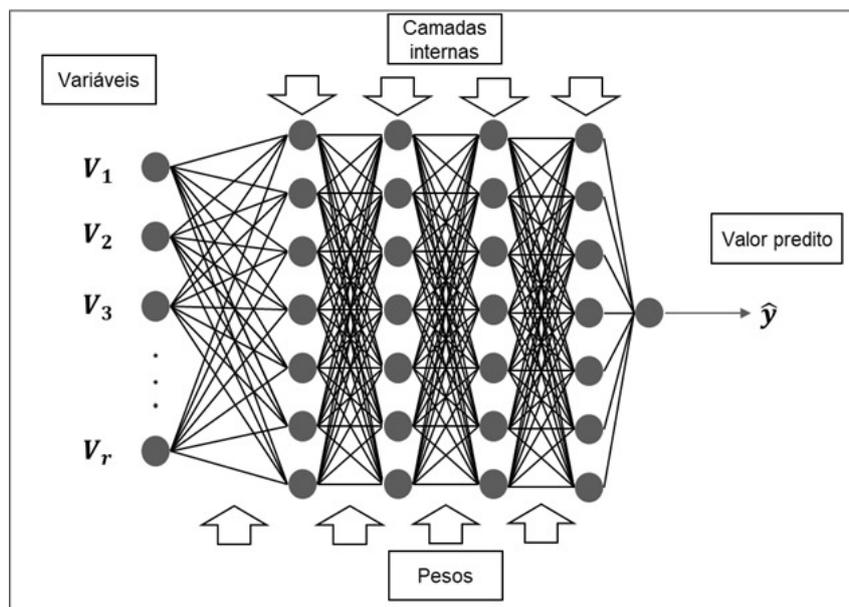


Figura 4 - Exemplo de arquitetura de rede neural expandida (com camadas internas)

Como forma de aproveitar o poder de otimização do Keras, além de comparar o GAMLSS com a Rede Neural Simples e o GAMLSS manualmente também ajustamos uma Rede Neural Profunda (expandida como a Figura 4) com 9 camadas internas e 1472 neurônios, ao todo são 350.466 parâmetros a serem estimados.

Em resumo, primeiramente criamos cinco diferentes modelos GAMLSS testando as variáveis explicativas. Todos utilizam a família da distribuição binomial negativa. Selecionamos um modelo GAMLSS candidato (“melhor GAMLSS”) e o comparamos com outros três tipos diferentes de modelo: 1) GAMLSS manualmente; 2) Rede Neural Simples (sem camadas inter-

nas) e; 3) Rede Neural expandida. Lembrando que esses três modelos utilizam funções de perda adaptadas da família da distribuição binomial negativa levando em conta a censura da variável resposta.

Para esta comparação, os quatro modelos foram treinados em uma base de treino utilizando 70% dos dados. Os quatro modelos foram comparados a partir da função de log-verossimilhança da distribuição binomial negativa com os respectivos valores preditos da base de teste (contendo os outros 30% dos dados).

## 5 Resultados

Para começar a apresentação dos resultados, foi escolhido um modelo aplicado nos dados de 2015 da base que faz o ajuste da média e variância e considera todas as variáveis não levando em conta as interações entre elas (modelo trivial - média e variância). A Tabela 6 apresenta os diagnósticos (incluindo os diagnósticos do modelo nulo, sem covariáveis). Os coeficientes estimados pra média e dispersão ( $\sigma$ ) estão na Tabela seguinte.

### 5.1 Modelos GAMLSS

	Modelo Trivial - média e variância	Modelo Nulo
Número de observações	20.162	20.162
Graus de liberdade do ajuste	50	2
Graus de liberdade dos resíduos	20.112	20.160
Desviância	94.073,28	96.837,4
Critério de Informação de Akaike	94.173,28	96.841,4
Critério de Informação Bayesiano	94.568,85	96.857,23

Tabela 6 - Diagnósticos do modelo trivial (média e variância) e nulo para o tempo até aposentadoria (em meses) em 2015.

Observa-se um ganho proporcional em relação ao Critério de Informação Bayesiano (BIC) de 2.36% do modelo trivial (média e variância) em relação ao modelo nulo. Pelos coeficientes

estimados no modelo trivial média e variância, os homens ficam, em média, 8.78% ( $e^{\beta_k} - 1$ , sendo  $\beta_k$  o coeficiente alvo estimado) mais tempo até se aposentar com todos os outros fatores constantes. O tempo de serviço público e a idade dos servidores apresentam padrões similares dos coeficientes estimados. A cada ano adicional vivido, espera-se, em média, que o servidor fique -2.3% meses até se aposentar.

Variáveis	Estimativa $\mu$	Erro padrão $\mu$	Estimativa $\sigma$	Erro padrão $\sigma$
Intercepto	4.2722	0.1428	-1.4363	0.2055
Gênero	0.0842	0.0157	-0.0337	0.0236
Tempo de serviço público	-0.0030	0.0013	0.0194	0.0019
Idade	-0.0232	0.0018	0.0024	0.0027
É professor	0.5587	0.0359	-0.2411	0.0537
Data de abono estimada	1.5159	0.0223	0.9595	0.0269
Salário líquido	0.0000	0.0000	0.0000	0.0000
Natureza jurídica - Autarquia Especial	-0.2398	0.0375	-0.0721	0.0583
Natureza jurídica - Autarquia Federal	-0.3124	0.0160	-0.2439	0.0258
Natureza jurídica - Fundação Federal	-0.2917	0.0223	-0.0388	0.0344
Região - Nordeste	0.1815	0.0212	-0.0796	0.0337
Região - Norte	0.3247	0.0235	-0.1237	0.0386
Região - Sudeste	-0.0738	0.0204	0.0785	0.0311
Região - Sul	-0.2250	0.0278	0.0972	0.0421
Estado Civil - Divorciado (a)	0.1208	0.0187	0.0396	0.0292
Estado Civil - Solteiro (a)	0.1798	0.0165	-0.0098	0.0268
Estado Civil - Viúvo (a)	-0.0366	0.0396	0.0396	0.0654
Escolaridade - Fundamental	0.1756	0.0486	0.1899	0.0788
Escolaridade - Médio	0.1875	0.0456	0.2314	0.0727
Escolaridade - Mestrado	-0.0211	0.0647	0.1640	0.0983
Escolaridade - Superior	-0.0123	0.0423	0.3035	0.0679
Jornada de trabalho (em horas)	0.0243	0.0016	0.0080	0.0024
Portador de deficiência física	-0.1009	0.1084	-0.0319	0.1456
Concursado - Sim	0.0545	0.0138	-0.1422	0.0223
Última classe da carreira	0.1024	0.0197	-0.2355	0.0316

Tabela 7 - Coeficientes estimados do modelo trivial (média e variância) para o tempo até aposentadoria (em meses) em 2015 (estimativas em log).

Há uma diferença notável entre as estimativas para os servidores da região norte e da região

sul. Os servidores que trabalham na região norte demoram em média 38.36% mais tempo em comparação com os servidores do centro-oeste, enquanto os servidores da região sul se aposentam -20.15% mais lentamente em relação ao centro-oeste. Por exemplo, um servidor do centro-oeste com valor médio estimado de 100 meses até a aposentadoria, considerando todos os mesmos fatores, a estimativa seria de 138.36 meses se estivesse na região norte, enquanto no sul, 79.85 meses.

Os servidores de natureza jurídica da administração direta federal, aqueles que integram Secretarias, Ministérios, Câmaras e Assembleias, são os que mais demoram para se aposentar em comparação com servidores de autarquias especiais ou federais ou fundações federais. É estimado que professores se aposentem com aproximadamente o dobro do tempo que um servidor não professor se aposenta (com todas as outras características iguais).

As amostras mais antigas, 2005, 2006 e 2007, mostram que parcelas consideráveis dos ativos aptos não optaram por se aposentar até 2019. Comportamento observado em todos os anos seguintes. Inclusive, a proporção de censuras é crescente ao longo do tempo. Os casos censurados na amostra de 2015 representam 56.2% dos casos (11.335 de 20.162), de acordo com a Tabela 4, o que impacta na estimativa dos casos não censurados. Observamos alguns exemplos de servidores com suas características e os respectivos valores estimados pelo modelo trivial (média e variância) na Tabela 8.

Apenas o exemplo 4 apresenta censura. O seu tempo observado até aposentadoria é de no mínimo 48 meses (aproximadamente o máximo observável da amostra de 2015 levando em conta a censura em julho de 2019). O valor estimado de 77 meses leva em conta esta censura. O exemplo 1 exemplifica a sobre estimativa observada na maioria dos casos em que a data de abono foi estimada – de acordo com o coeficiente estimado da variável Data de abono estimada, aqueles que tiveram esta estimativa se aposentam (tudo o mais constante) 355.36% mais tarde. Os exemplos 2 e 3, por outro lado, apresentam valores estimados próximos dos observados. O modelo nulo que apresenta apenas intercepto ajusta todos os casos para o mesmo valor em 86 meses, o que não leva em conta as particularidades e características específicas de cada servidor.

Variável	Exemplo 1	Exemplo 2	Exemplo 3	Exemplo 4
Censura	Não	Não	Não	Sim
Concursado	Não	Sim	Sim	Não
Data de abono estimada	Sim	Não	Não	Não
Escolaridade	Médio	Superior	Superior	Superior
Estado Civil	Solteiro	Divorciado	Solteiro	Casado
Gênero	Mulher	Homem	Mulher	Homem
Idade	54	59	52	57
Jornada de trabalho	40	20	40	40
Natureza jurídica	Administração Direta	Administração Direta	Autarquia Federal	Administração Direta
Portador de deficiência física	Não	Não	Não	Não
Professor	Não	Não	Não	Não
Região	Centro Oeste	Sudeste	Sudeste	Sudeste
Salário líquido	3395	9202	5356	12692
Servidor na última classe da carreira	Sim	Sim	Sim	Sim
Tempo de serviço público	28	30	26	36
Meses até idade compulsória	240	192	264	203
Meses até aposentadoria	19	46	50	48+
Meses estimados pelo Modelo	402	49	58	77
Meses estimados Modelo Nulo	86	86	86	86

Tabela 8 - Exemplos de servidores e as estimativas pelo modelo trivial media e variancia (Modelo) e Modelo Nulo em 2015.

A análise desenvolvida não se restringe apenas ao modelo ajustado para amostra de 2015. O fato é que há um enorme conjunto de modelos. Não cabe interpretar os resultados dos coeficientes estimados de cada modelo em cada ano nesta Seção.

Até aqui, a ênfase foi apenas no modelo trivial (média e variância) em comparação com o modelo nulo. Acreditamos há outros dois conjuntos de modelos que sejam mais complexos, ao passo que explicam apenas levemente melhor os dados observados. Observamos, a partir da Figura 5, nos anos de 2005 a 2013 da amostra, que o modelo trivial (apenas média) apresenta um ganho proporcional de aproximadamente metade do ganho do modelo mais complexo (modelo com interação e ajuste para média e variância). Enquanto que os modelos menos complexos – trivial e trivial stepwise que também modela a média e variância – têm ganhos proporcionais

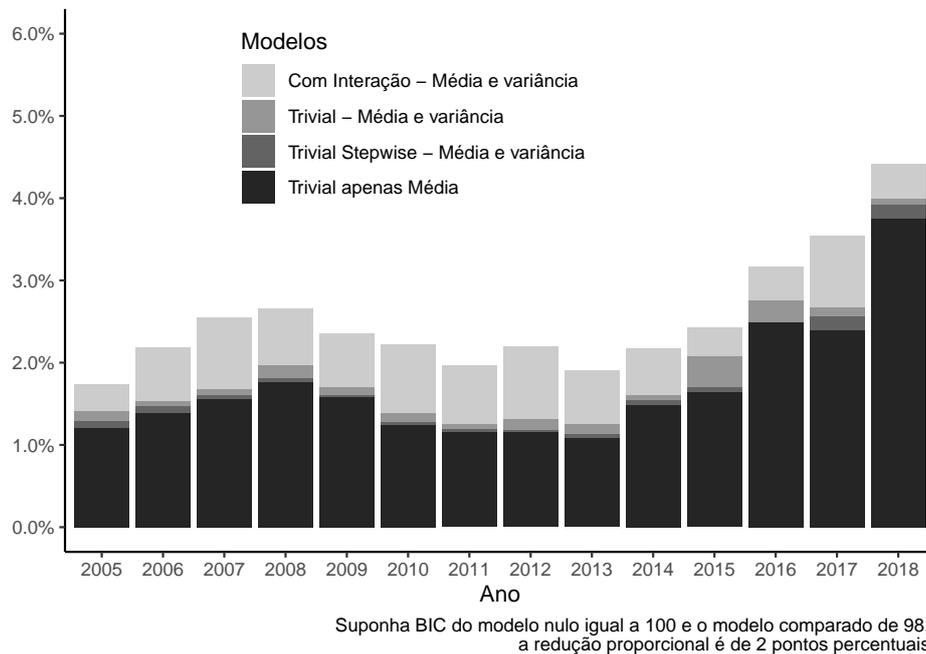


Figura 5 - Comparação entre os ganhos proporcionais das medidas de BIC dos diferentes modelos para cada ano em relação ao modelo nulo.

com diferenças imperceptíveis entre si e são muito próximos do modelo mais complexo. O que representa uma vantagem por serem modelos mais parcimoniosos ao passo que explicam bem os dados. Os modelos de anos mais recentes, 2016 a 2018, apresentam os melhores resultados em relação ao modelo nulo.

Devido aos ganhos relativos entre os modelos com variáveis e o modelo nulo serem muito próximos, opta-se por utilizar o modelo trivial (média e variância) que utiliza o maior número de variáveis em relação ao modelo trivial stepwise (média e variância) ao passo que continua parcimonioso (ausência de interações como o modelo “Com interação stepwise para a média e variância”). Ou seja, melhor interpretabilidade. A partir deste ponto do texto, utilizaremos o modelo GAMLSS com todas as variáveis para média e dispersão como o modelo escolhido para comparação com outros modelos.

## 5.2 Modelos alternativos

Propomos desenvolver três modelos alternativos comparáveis para examinar quão bem os modelos GAMLSS, escolhidos na Seção anterior, explicam os dados. Estes métodos alternativos utilizam as mesmas variáveis do modelo GAMLSS para estimar – todas as variáveis como na Tabela 7 – a média e a dispersão.

Por terem arquiteturas diferentes, especialmente os dois tipos de redes neurais testadas, estes modelos poderiam capturar relações entre as variáveis que os modelos lineares não capturariam. O que é uma vantagem comparativa para ser testada. O primeiro modelo alternativo é chamado GAMLSS manualmente, explicado na Seção 4.3, em que propomos estimar os coeficientes do modelo minimizando a função de log-verossimilhança negativa desenvolvida a partir de uma função de otimização do software R. O segundo modelo é uma rede neural com arquitetura simples. Não apresenta camadas internas. Liga as variáveis diretamente com os valores preditos para a média e a variância. O que pode ser considerado como uma implementação de um modelo de regressão binomial negativa utilizando o processo de otimização de rede neural. O terceiro modelo, entretanto é mais complexo em sua arquitetura. Foram ajustadas 9 camadas internas com respectivamente 32, 64, 128, 256, 512, 256, 128, 64 e 32 neurônios em cada camada. Ao todo são 1472 neurônios e 350.466 parâmetros a serem treinados – o que dá a ordem de complexidade deste modelo.

Como mencionado, os três modelos alternativos são treinados a partir de uma função de perda para adaptar a função de log verossimilhança da distribuição binomial negativa tratando a censura para estimar a média e a dispersão de cada observação. Utilizamos o valor dessa função adaptada para comparar os quatro métodos ano a ano – observado na Figura 6. Quanto maior seu valor, melhor o modelo. O gráfico apresenta a diferença percentual entre cada modelo alternativo e o modelo GAMLSS alvo, escolhido na Seção anterior. Por exemplo, se o valor da função de log-verossimilhança do GAMLSS é igual a 100 e algum modelo alternativo é igual a 102, temos uma diferença percentual de 2%.

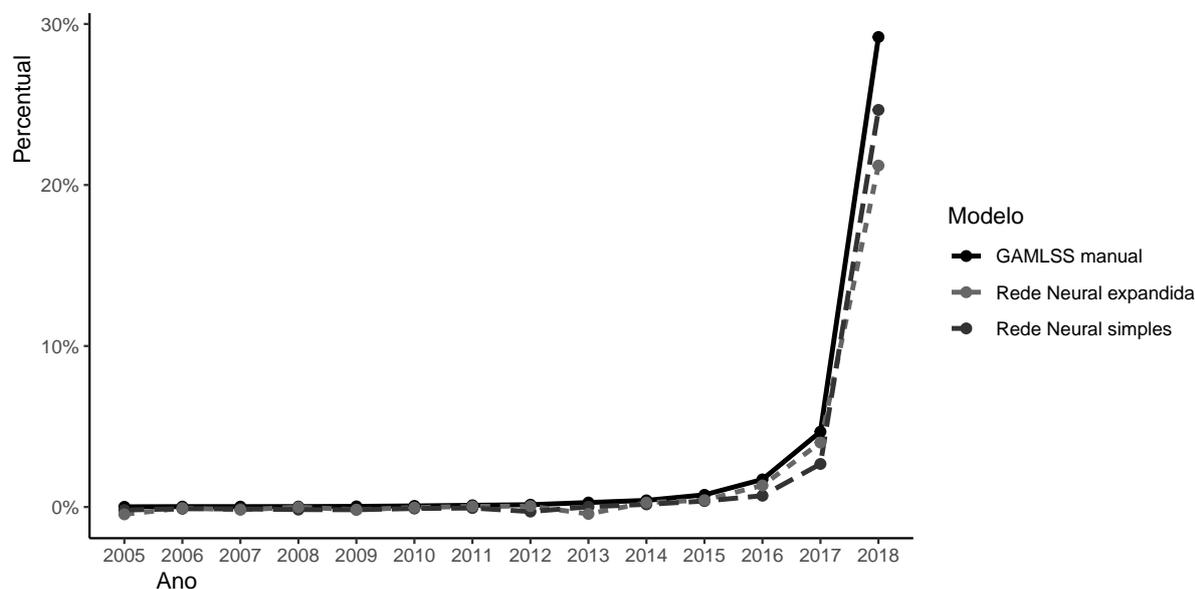


Figura 6 - Diferença percentual entre log-verossimilhança de cada método de otimização e do GAMLSS por ano.

Observamos que os modelos alternativos dos anos de 2005 a 2016 não apresentam ganhos significativos em relação ao modelo GAMLSS. A maior diferença percentual observada neste período é de 1,7% do modelo GAMLSS manualmente em 2016. Em geral, observa-se diferenças bem próximas a zero. As diferenças percentuais de 2017 são maiores – na ordem de 4%. Por fim, destaca-se um ganho significativo em 2018. Vale pontuar que nos anos mais recentes a maioria dos dados são casos censurados. Como visto na Tabela 4, 71% dos dados são censurados em 2017 e 80% em 2018. O que representa um desbalanceamento nos dados.

Portanto, concluí-se que os modelos GAMLSS proposto conseguem capturar bem as informações contidas nas observações para explicar os tempos até a aposentadoria. As diferenças percentuais próximas de zero na maioria dos anos fortalece a decisão de aplicação dos modelos GAMLSS, principalmente em detrimento da interpretabilidade dos coeficientes estimados – Como feito na Seção 5.1.

### 5.3 Modelos da literatura

Por fim, o modelo proposto é comparado qualitativamente com outros modelos significativos da literatura. Vale aprofundar em quatro textos já mencionados na Seção 1.1 para debater sobre cada um em comparação com o modelo proposto. São eles: Schettini, Santos e Pires (2018), Santos, Pires e Mendonça (2020), Salazar e Penas (2019) e Samwick (1998)

O primeiro texto utiliza regras previdenciárias para ter uma estimativa de quando os servidores estarão aptos a se aposentarem. A partir deste ponto do servidor, para simular o padrão histórico de aposentadorias passadas, o trabalho assume que os riscos expirados (servidores que já adquiriram direito a aposentadoria) têm uma probabilidade de aposentar de acordo com a idade do indivíduo normalizada,  $x = (y - y_{min}) / (y_{max} - y_{min})$  (no qual  $x$  é a idade normalizada,  $y$  é a observada) usando uma distribuição de probabilidade Beta com parâmetros 4 e 2.

A ideia desenvolvida em Schettini, Santos e Pires (2018) para quando os servidores se aposentarão funciona como um insumo (ou hipótese) do propósito do artigo, que é simular gastos com aposentadoria caso uma PEC fosse promulgada. Entretanto, como modelagem é simples em seu desenho. Utiliza-se apenas da idade.

O segundo texto também utiliza informações pessoais dos servidores para estimar, com base nas regras previdenciárias, quando o direito a aposentadoria é adquirida. Entretanto, é adicionada mais complexidade: utiliza mais variáveis (tempo apto a aposentadoria e cargo) além de ter um modelo mais robusto. Diferentemente do primeiro artigo, os autores modelaram a saída dos servidores a partir do momento que adquire o direito a aposentadoria com um modelo de análise de sobrevivência para quatro tipos de cargo de servidor – professores, policiais civis ou agentes penitenciários, militares ou outros estatutários. Neste modelo, os servidores deixam o serviço público para a inatividade de acordo com o tempo apto a aposentadoria e o seu cargo. Quanto mais tempo apto a se aposentar, maior é a probabilidade de saída.

Neste caso, o modelo desenvolvido para estimar a aposentadoria também se propõe a ser

um insumo para um modelo de simulação. Neste caso, uma projeção de gastos com pessoal do serviço público estadual.

Em comparação com estes dois artigos, o modelo proposto neste texto viabiliza a utilização de uma granularidade maior de informações acerca de cada indivíduo. Apesar de não ser o foco dos artigos mencionados, aqui fazemos uma comparação mais robusta entre diferentes técnicas de estimação.

Os outros dois artigos, Salazar e Penas (2019) e Samwick (1998), empregam técnicas estatísticas e de aprendizagem de máquinas. O primeiro artigo utiliza dados de previdência privada do México para prever se os indivíduos se aposentam antes ou após atingir 65 anos de idade em função de características pessoais dos indivíduos como gênero, se a pessoa possui alguma condição física relacionada às 10 maiores causas de mortalidade no país, grau de educação, estado civil, salário, status de emprego, número de dependentes, score de crédito, e algumas variáveis macroeconômicas como retorno de títulos do governo, taxa de desemprego e um índice do mercado de ações.

São testados modelos de classificação como Regressão Logística, *Support Vector Machine* e *Random Forest*, o que dá robustez ao trabalho. Entretanto, a amostra é composta apenas daqueles que já aposentaram. Não são incluídos observações em que o indivíduo já adquiriu o direito a aposentadoria mas ainda não optou por fazê-lo.

O segundo artigo, Samwick (1998), utiliza dados de uma pesquisa sobre o histórico de aposentadoria americana (RHS – *Retirement Historic Survey*). A pesquisa é um painel de 1969 contendo informações bienais de 10 anos de 8 mil homens com idades entre 58 e 63 anos. As informações são sobre trabalho, salário, cobertura de previdência e estado de saúde. O estudo tenta prever se um indivíduo optou por se aposentar ou não usando um modelo Probit (Modelo linear generalizado para respostas binárias que utiliza a função de ligação probit) e conclui que o acúmulo de previdência é um fator de extrema importância para esta previsão, inclusive se o modelo previdenciário é alterado.

A relevância das variáveis utilizadas é notável neste estudo por levar em conta informações

Variável	Modelo proposto	Schettini, Santos e Pires (2018)	Santos, Pires e Mendonça (2020)	Salazar e Penas (2019)	Samwick (1998)
Censura	✓				
Concurado	✓				
Data de abono estimada	✓				
Escolaridade	✓			✓	✓
Estado Civil	✓			✓	✓
Gênero	✓			✓	✓
Idade	✓	✓			✓
Jornada de trabalho	✓				
Natureza jurídica	✓				
Portador de deficiência física	✓				
Professor	✓		✓		
Região	✓				
Salário líquido	✓			✓	✓
Servidor na última classe da carreira	✓				
Tempo de serviço público	✓				
Cargo			✓		✓
Condição de saúde				✓	✓
Status empregador/empregado				✓	
Número dependentes				✓	
Taxa de desemprego				✓	
Score de crédito				✓	
Retorno de títulos públicos				✓	
Índice ações				✓	
Acúmulo de previdência					✓
Não branco					✓
Recursos financeiros					✓
Patrimônio líquido (habitação)					✓

Tabela 9 - Resumo de variáveis utilizadas em cada artigo comparado.

previdenciárias individuais (planos de previdência e o respectivo acúmulo financeiro) e composição familiar. A disponibilidade de quanto cada indivíduo acumulou durante sua carreira naturalmente é um fator decisivo na decisão de se aposentar. Apesar disso, o texto debate sobre a amostra utilizada – homens entre 58 e 63 anos – que não é representativa da classe trabalhadora de forma geral. Vale mencionar que este texto se propõe a generalizar os resultados

encontrados em Stock e Wise (1990) que utilizam uma amostra de trabalhadores de apenas uma empresa (com apenas um plano de previdência) para uma amostra com mais tipos de planos de previdência.

	Modelo proposto	Schettini, Santos e Pires (2018)	Santos, Pires e Mendonça (2020)	Salazar e Penas (2019)	Samwick (1998)
Variável resposta	Tempo até aposentadoria	Probabilidade de aposentadoria	Probabilidade de aposentadoria	Aposentou antes ou depois de 65 anos	Aposentou ou não
Técnica utilizada	GAMLSS	Distribuição Beta	Análise de sobrevivência	Random Forest, Regressão Logística e Support Vector Machine	Modelo Linear Generalizado

Tabela 10 - Resumo de técnicas utilizadas e desenho da variável resposta em cada artigo comparado.

Em relação ao proposto neste estudo e os artigos de Samwick (1998) e Salazar e Penas (2019) uma grande diferença está no desenho da variável resposta. Os dois artigos a utilizam com uma visão binária – aposentou ou não. Enquanto, propomos estimar o tempo até a aposentadoria além de levar em conta indivíduos que estão aptos a se aposentar mas ainda não o fizeram por escolha. O que dá um escopo maior ao estudo. Entretanto, vale ressaltar variáveis importantes utilizadas em cada um dos métodos. O primeiro artigo leva em conta fatores estruturais macroeconômicos enquanto o segundo artigo utiliza de informações acerca do acúmulo de capital previdenciário dos indivíduos. A Tabela 9 resume as variáveis utilizadas e a Tabela 10 as técnicas utilizadas em cada estudo comparado.

## 5.4 Suavização dos coeficientes estimados pelos modelos GAMLSS

Como opção de modelagem, descrito na Seção 4, foi decidido por modelar cada ano separadamente. Dessa forma, captura-se os efeitos de cada característica do indivíduo separadamente por ano. Foge do escopo deste trabalho criar um novo modelo para prever esses efeitos. Porém, propomos um meta modelo (Figuras 7 e 8) de suavização com o método LOESS (*Locally Estimated Scatterplot Smoothing*), técnica não paramétrica de regressão proposta inicialmente por

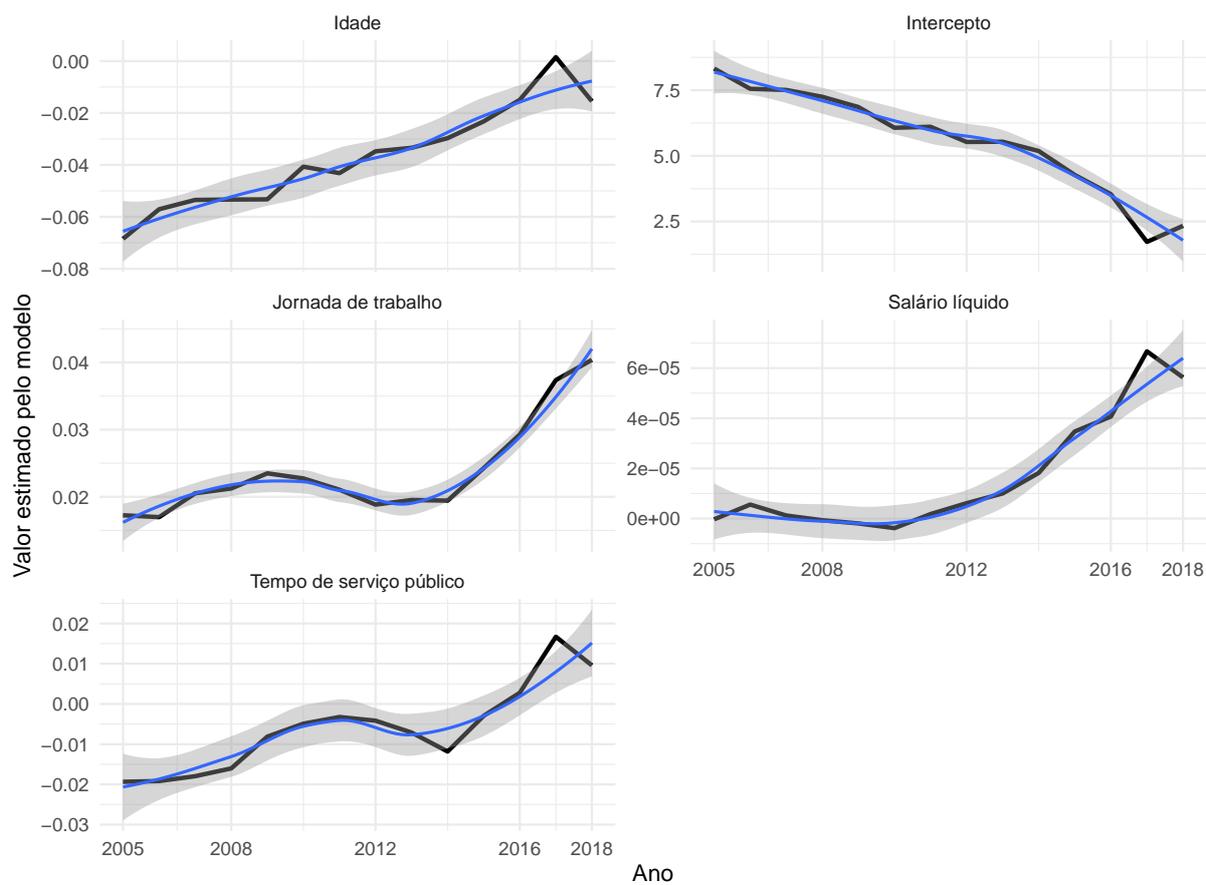


Figura 7 - Coeficientes estimados (em preto) para média das variáveis numéricas dos modelos GAMLSS ano a ano com suavização (em azul) e erro padrão (em cinza)

Cleveland (1979). Método que não precisa atender suposições da relação entre as variáveis.

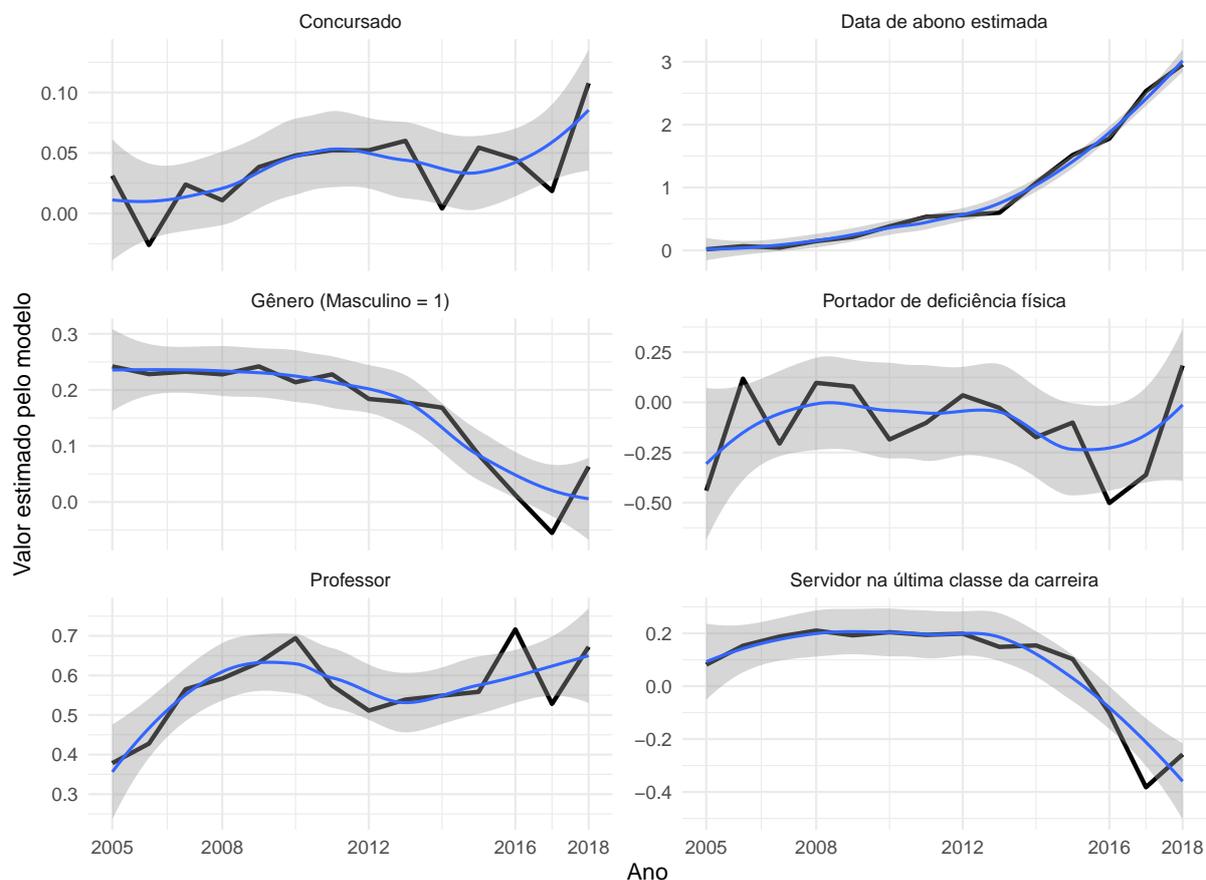


Figura 8 - Coeficientes estimados (em preto) para média das variáveis categóricas binárias dos modelos GAMLSS ano a ano com suavização (em azul) e erro padrão (em cinza)

Essa técnica nos permite observar a evolução das estimativas dos coeficientes sem as variações anuais individuais. Assim observa-se as tendências ao longo do tempo do efeito de cada variável nos modelos GAMLSS.

Observamos, na Figura 7, por exemplo, que a variável Idade vai diminuindo o seu efeito ao longo do tempo. No modelo de 2005, com todos os outros fatores constantes a cada ano adicional de um servidor equivalia a -6% meses até a aposentadoria na estimação final. O efeito do salário líquido vai aumentando ao longo do tempo também. Indicando que o efeito de um mesmo salário vai aumentando ao longo dos anos.

O efeito do gênero (Figura 8), entretanto, possui a relação inversa. Ao longo dos anos, o impacto da variável diminui nas estimativas de tempo até a aposentadoria. Assim como a variável indicadora do Servidor em última classe de carreira. Os gráficos de tendência das estimativas de coeficiente para média das outras variáveis estão no Apêndice nas Figuras 9, 10, 11, 12.

## 5.5 Truncamento dos valores estimados pelo modelo GAMLSS

Até agora, a idade compulsória de aposentadoria do servidor federal não foi incorporada nos modelos desenvolvidos. É fato que, pela lei previdenciária, o servidor é aposentado compulsoriamente quando atinge 75 anos de idade. Propomos então, tratar os valores estimados para que essa regra previdenciária seja levada em conta. Ou seja, para que não haja estimativas de tempo até a aposentadoria maiores que o tempo até a idade compulsória do servidor.

Utilizamos a função *dtrunc* do pacote *truncdist*, Nadarajah e Kotz (2006), do software R que permite calcular a densidade truncada da binomial negativa a partir dos coeficientes de média e dispersão estimados em cada modelo e o truncamento (idade até a compulsória). Com isso, reajustamos o tempo até a aposentadoria média estimada com a média truncada, a partir desta função, como a equação a seguir.

$$\hat{\mu}_{trunc} = \sum_{x=0}^k x f_{trunc}(x|\hat{\mu}, \hat{\sigma}, k)$$

Em que  $\hat{\mu}_{trunc}$  é a média estimada até a aposentadoria tratada com a Idade Compulsória,  $k$  é o tempo até a idade compulsória (truncamento),  $\hat{\mu}$  é a média ajustada pelo modelo,  $\hat{\sigma}$  é a dispersão ajustada pelo modelo para este exemplo e  $f_{trunc}$  é a função densidade truncada da binomial negativa.

Por termos modelado a média e a dispersão com o GAMLSS, temos estimativas para estes dois coeficientes para cada observação. A partir destes valores, calculamos, com a Função de Distribuição Acumulada da Binomial Negativa,  $F(X > a) = 1 - F(X \leq a)$  (em que  $a$  é o Tempo até a Idade Compulsória), a probabilidade estimada de aposentadoria após a Idade

Variável	Exemplo 1	Exemplo 2	Exemplo 3	Exemplo 4
Censura	Não	Não	Não	Sim
Concurado	Não	Sim	Sim	Não
Data de abono estimada	Sim	Não	Não	Não
Escolaridade	Médio	Superior	Superior	Superior
Estado Civil	Solteiro	Divorciado	Solteiro	Casado
Gênero	Mulher	Homem	Mulher	Homem
Idade	54	59	52	57
Jornada de trabalho	40	20	40	40
Natureza jurídica	Administração Direta	Administração Direta	Autarquia Federal	Administração Direta
Portador de deficiência física	Não	Não	Não	Não
Professor	Não	Não	Não	Não
Região	Centro Oeste	Sudeste	Sudeste	Sudeste
Salário líquido	3395	9202	5356	12692
Servidor na última classe da carreira	Sim	Sim	Sim	Sim
Tempo de serviço público	28	30	26	36
Meses até idade compulsória	240	192	264	203
Meses até aposentadoria	19	46	50	48+
Meses estimados pelo Modelo	402	49	58	77
<b>Meses pelo Modelo truncado</b>	<b>139</b>	<b>43</b>	<b>50</b>	<b>66</b>
Meses estimados Modelo Nulo	86	86	86	86

Tabela 11 - Exemplos de servidores e as estimativas pelo Modelo trivial media e variancia (Modelo), Modelo trivial tratado com truncamento e Modelo Nulo em 2015.

Compulsória. Observamos que em muitos casos essa probabilidade é maior que 0. Até mesmo, algumas estimativas de tempo até a aposentadoria superam o Tempo até a Idade Compulsória. Por exemplo, o modelo ajustado pelo GAMLSS em 2015 tem aproximadamente 17.62% (1066 casos) com estimativas do tempo até a aposentadoria maiores que o Tempo até a Idade Compulsória. O exemplo 1 da Tabela 11 (Tabela 8 adicionando o tratamento da estimativa inicial) é um desses casos. O tempo até aposentadoria é de 19 meses (não censurado) e até a idade compulsória é de 240 meses. O modelo estimou inicialmente 402 meses, o que ultrapassaria a idade compulsória. Após o tratamento, obtemos a estimativa de 138 meses.

Após este truncamento, não se observa nos modelos de nenhum ano casos de estimativas da

idade de aposentadoria maiores que a Idade compulsória e probabilidades maiores que zero em valores maiores que o tempo até a idade compulsória.

## 6 Considerações finais

Esta dissertação apresentou modelos estocásticos que tentam explicar os fatores que levam os servidores da União a se aposentarem. Isso só foi possível pela disponibilidade da base de dados bruta fornecida pela Secretária de Previdência. Base com formato fundamental e decisivo para a análise proposta. Além de toda a metodologia proposta para ajustar esta base bruta para modelagem descrita nas Seções 3 e 4.

Foram ajustados vários tipos de modelo por ano utilizando a técnica estatística GAMLSS (*Generalized Additive Model for Location Scale and Shape*) com diferentes variáveis para comparação. Foram escolhidos os modelos em que foi utilizado todas as variáveis disponíveis para ajustar a média e a dispersão, coeficientes da família binomial negativa. Esta decisão foi tomada com base na parcimônia dos dados utilizados e na qualidade do ajuste em comparação com os outros modelos testados.

Este modelo escolhido foi então colocado à prova comparando-os com Redes Neurais utilizando uma função de perda adaptada para tornar os dois tipos de modelos comparáveis. Duas arquiteturas diferentes de redes neurais foram utilizadas. Concluiu-se que os modelos GAMLSS explicam bem os dados e não há evidências para utilização das redes neurais em detrimento do modelo já ajustado. Além disso, a metodologia aplicada foi comparada qualitativamente com outras quatro abordagens em relação as técnicas utilizadas, escolhas de modelagem e variáveis utilizadas.

Os coeficientes dos modelos GAMLSS ajustados de cada ano foram suavizados com uma técnica de regressão não paramétrica, LOESS (*Locally Estimated Scatterplot Smoothing*) e os valores estimados do tempo até a aposentadoria foram tratados para incorporar a regra previdenciária de aposentadoria em idade compulsória.

Em conclusão, a função de perda levando em conta os dados censurados, as redes neurais e o tratamento do truncamento devido a idade compulsória de aposentadoria são os principais desenvolvimentos técnicos propostos por este texto. Estes pontos e outros desenvolvimentos tratados ao longo do texto tentam solucionar os quatro desafios mencionados na Seção 1.2 – 1) Censuras; 2) Idade compulsória; 3) tempo de trabalho anterior e; 4) dados de boa qualidade.

## Referências

- ALLAIRE, J.; CHOLLET, F. **keras: R Interface to 'Keras'**. [s.l: s.n.].
- BHERING, J. W. S. Simulação Estocástica em Plano de Contribuição Definida. **UFRJ/IM**, 2005.
- BROWN, T. H.; WARNER, D. F. Divergent pathways? Racial/ethnic differences in older women's labor force withdrawal. **Journals of Gerontology: Social Sciences**, v. 63B, n. 3, p. S122–134, 2008.
- BYRD, R. H. *et al.* A limited memory algorithm for bound constrained optimization. **SIAM Journal on Scientific Computing**, v. 16, p. 1190–1208, 1995.
- CASALECCHI, A. R. C.; ORAIR, R. O.; SOUZA, P. H. O. Simulação do Impacto da PEC 6/2019 sobre as despesas do RPPS dos Servidores Civis da União. **Caderno de Finanças Públicas - "Edição Especial"**, v. 1, n. 1, 2020.
- CHOUDHURY, S.; BRIDGES, B. Examining Social Security benefits as a retirement resource for near-retirees, by race and ethnicity, nativity and disability status. **Social Security Bulletin**, v. 69, n. 1, p. 19–44, 2009.
- CLEVELAND, W. S. Robust Locally Weighted Regression and Smoothing Scatterplots. **Journal of the American Statistical Association**, v. 74, p. 829–836, 1979.
- DEKKERS, G. *et al.* On using dynamic microsimulation models to assess the consequences of the AWG projections and hypotheses on pension adequacy: Simulation results for Belgium, Sweden and Hungary. **Relatório técnico da agência belga Federal Planning Bureau**, mar. 2015.
- DOWLE, M.; SRINIVASAN, A. **data.table: Extension of 'data.frame'**. [s.l: s.n.].
- FREDRIKSEN, D. Projections of Population, Education, Labour Supply and Public Pension Benefits: Analyses with the dynamic microsimulation model MOSART. **Statistisk sentralbyrå**, jan. 1998.
- GROLEMUND, G.; WICKHAM, H. Dates and Times Made Easy with lubridate. **Journal of**

**Statistical Software**, v. 40, n. 3, p. 1–25, 2011.

KENC, T. Modelling The Retirement Decision: A Martingale Duality Approach. fev. 2000.

KIM, H.; DEVANEY, S. A. The selection of partial or full retirement by older workers. **Journal of Family and Economic Issues**, v. 26, n. 3, p. 371–394, 2005.

MAESTAS, N. Back to work: expectations and realizations of work after retirement. **Journal of Human Resources**, v. 45, n. 3, p. 718–748, 2010.

MUTCHLER, J. E. *et al.* Pathways to labor force exit: work transitions and work instability. **Journals of Gerontology: Social Sciences**, v. 52B, n. 1, p. S4–12, 1997.

MUTCHLER, J. E. *et al.* Work transitions and health in later life. **Journals of Gerontology: Social Sciences**, v. 54B, n. 5, p. S252–261, 1999.

NADARAJAH, S.; KOTZ, S. R Programs for Computing Truncated Distributions. **Journal Statistical Software**, v. 16(2), p. 1–8, 2006.

NELSON, W. **Censored Data and Maximum Likelihood Methods**. [s.l.] Wiley Series in Probability; Statistics, 1990.

PIENTA, A. M.; BURR, J. A.; MUNTCHLER, J. E. Women’s labor force participation in later life: the effects of early work and family experience. **Journals of Gerontology: Social Sciences**, v. 49, n. 5, p. S231–239, 1994.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing, 2021.

RIGBY, R. A.; STASINOPOULOS, D. M. The GAMLSS project: a flexible approach to statistical modelling. **Klein, B. and Korsholm, L., New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical Modelling**, p. 249–256, 2001.

\_\_\_\_. Generalized additive models for location, scale and shape, with discussion. **Applied Statistics**, v. 54, p. 507–554, 2005.

\_\_\_\_. Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. **Journal of Statistical Software, Articles**, v. 23, n. 7, p. 1–46, 2007.

ROGOWSKI, J. A.; KAROLY, L. Health insurance and retirement behavior: evidence from the

Health and Retirement Survey. **Journal of Health Economics**, v. 19, n. 4, p. 529–539, 2000.

SALAZAR, J. J. R.; PENAS, M. C. B. Scoring and prediction of early retirement using machine learning techniques: application to pension plans. **Anales del Instituto de Actuarios Españoles**, v. 4, n. 25, p. 119–145, 2019.

SAMWICK, A. A. New evidence on Pensions, Social Security, and the Timing of Retirement. n. 6534, 1998.

SCHETTINI, B. P.; PIRES, G. M. V. Gastos com pessoal no contexto do Novo Regime Fiscal: Algumas simulações para o Poder Executivo civil. **Carta de Conjuntura do IPEA**, v. 41, 2018.

SCHETTINI, B. P.; SANTOS, C. H. M.; PIRES, G. M. V. Previdência e reposição no serviço público civil federal do poder executivo: microssimulações. **Texto para discussão do IPEA**, v. 2365, 2018.

STOCK, J. H.; WISE, D. A. Pensions, the option value of work, and retirement. **Econometrica**, v. 58, n. 5, p. 1151–1180, 1990.

TANG, F.; BURR, J. A. Revisiting the pathways to retirement: a latent structure model of the dynamics of transition from work to retirement. **Ageing and Society**, p. 1–32, jun. 2014.

WARNER, D.; HAYWARD, M.; HARDY, M. The retirement life course at the end of the 21st century. **Population Research and Policy Review**, v. 29, p. 893–919, 2010.

WICKHAM, H. *et al.* Welcome to the tidyverse. **Journal of Open Source Software**, v. 4, n. 43, p. 1686, 2019.

WONG, J. D.; HARDY, M. A. Women’s retirements expectations: how stable are they? **Journals of Gerontology: Social Sciences**, v. 64B, n. 1, p. 77–86, 2009.

WRIGHT, R. Paying for retirement: sex differences in inclusion in employer-provided retirement plans. **The Gerontologist**, v. 52, n. 2, p. 231–244, 2012.

## Apêndice

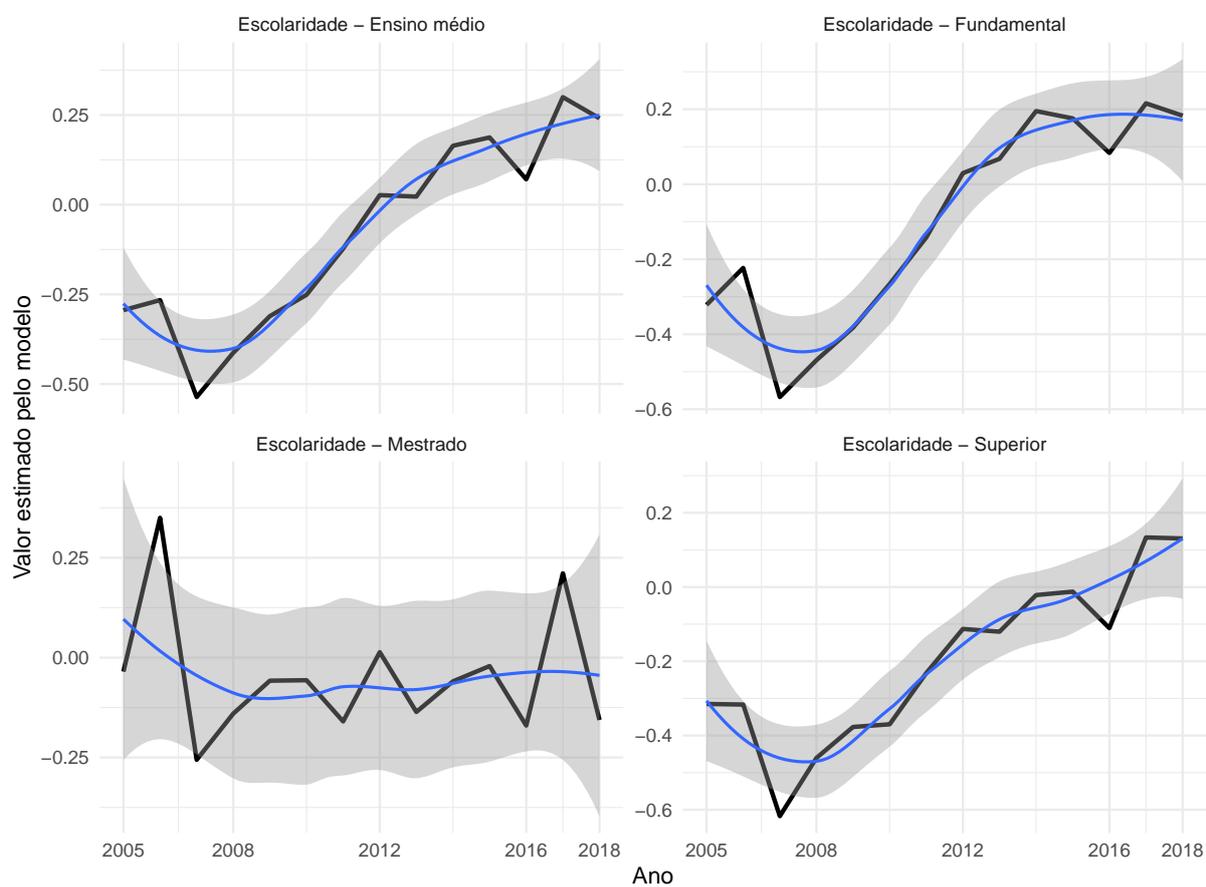


Figura 9 - Coeficientes estimados (em preto) das variáveis de escolaridade dos modelos GAMLSS ano a ano com suavização (em azul) e erro padrão (em cinza)



Figura 10 - Coeficientes estimados (em preto) para média das variáveis de estado civil dos modelos GAMLSS ano a ano com suavização (em azul) e erro padrão (em cinza)

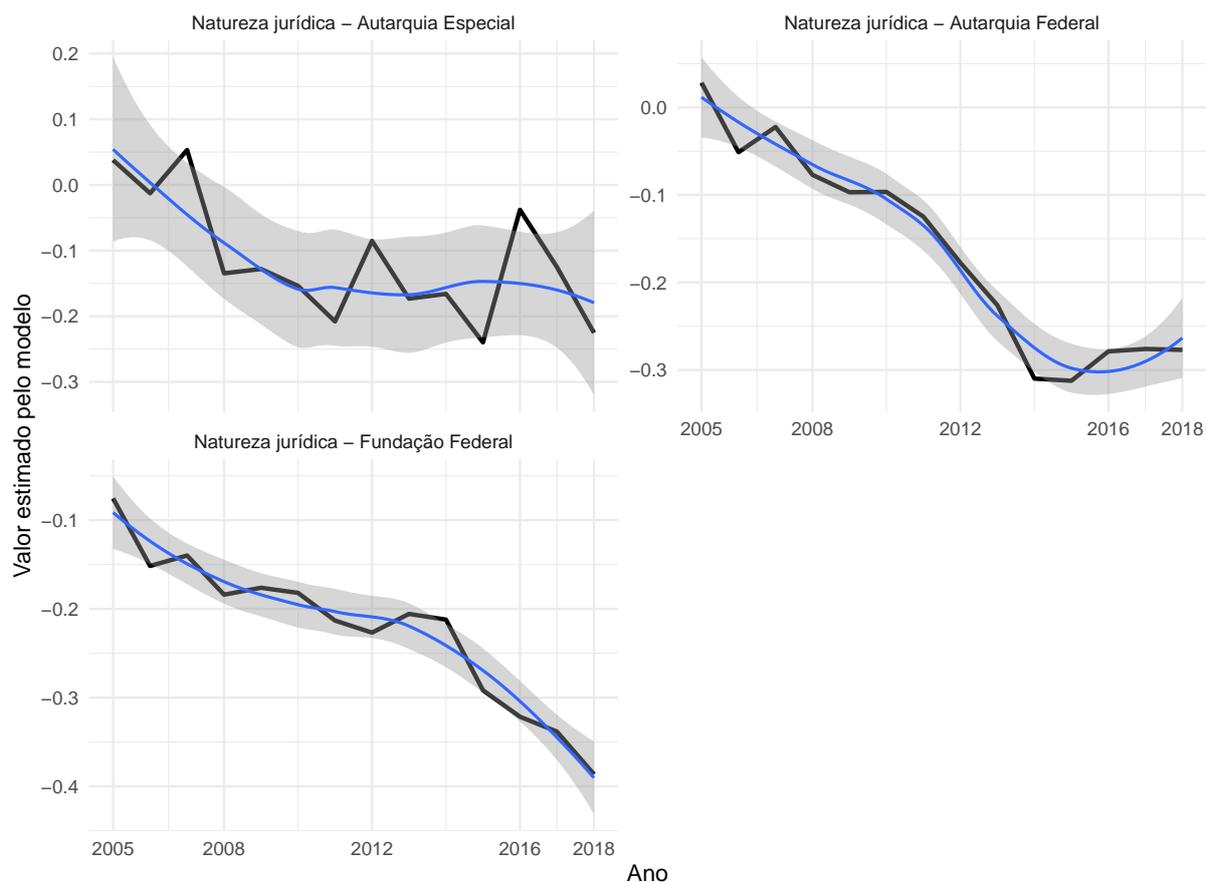


Figura 11 - Coeficientes estimados (em preto) para média das variáveis de natureza jurídica dos modelos GAMLSS ano a ano com suavização (em azul) e erro padrão (em cinza)

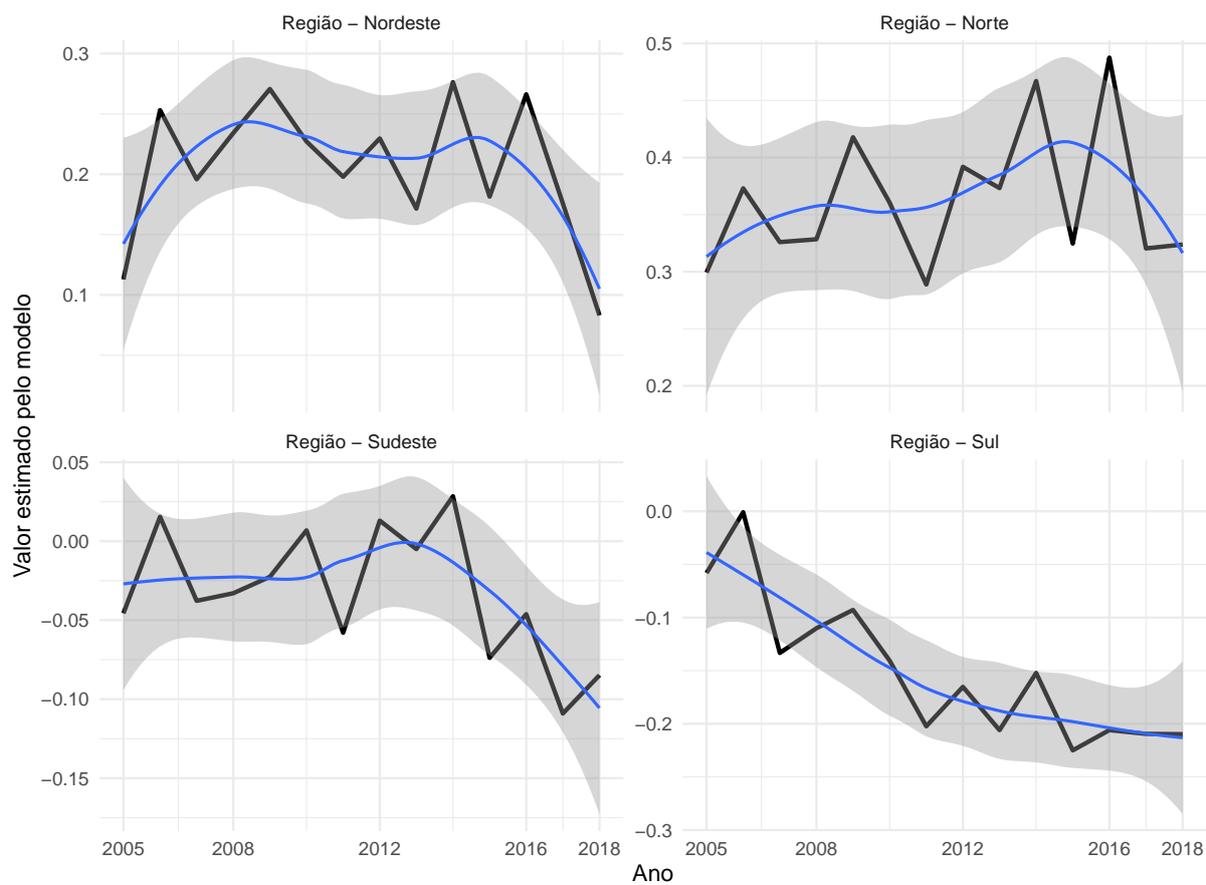


Figura 12 - Coeficientes estimados (em preto) para média das variáveis de região dos modelos GAMLSS ano a ano com suavização (em azul) e erro padrão (em cinza)