

# Métodos de previsão de prêmios para o Seguro Agrícola e destinação de recursos públicos ao Programa de Subvenção ao Prêmio do Seguro Rural

*Premium forecasting methods for Crop Insurance and allocation of federal resources to the Crop Insurance Premium Subsidy Program*

Arthur Augusto Lula Mota<sup>1</sup> , Vitor Ozaki<sup>1</sup> , Daniel Lima Miquelluti<sup>1</sup> 

<sup>1</sup>Departamento de Economia, Administração e Sociologia, Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ), Universidade de São Paulo (USP), Piracicaba (SP), Brasil. E-mails: arthur.lula.mota@gmail.com; vitorozaki@usp.br; danielmiq@usp.br

**Como citar:** Mota, A. A. L., Ozaki, V., & Miquelluti, D. L. (2022). Métodos de previsão de prêmios para o Seguro Agrícola e destinação de recursos públicos ao Programa de Subvenção ao Prêmio do Seguro Rural. *Revista de Economia e Sociologia Rural*, 60(spe), e249013. <https://doi.org/10.1590/1806-9479.2021.249013>

**Resumo:** A previsão dos prêmios do seguro agrícola tem impacto relevante no dimensionamento orçamentário do Programa de Subvenção ao Prêmio do Seguro Rural (PSR). Pela ótica da responsabilidade e do planejamento do gasto público, é preciso estimar com relativa precisão quanto se pretende despende com o Programa no curto e médio prazo (Emenda Constitucional nº 95, de 2016). O presente estudo faz uma comparação de métodos para a projeção do prêmio do seguro agrícola por região. Foram utilizados os modelos SARIMA e os algoritmos NNAR, TBATS, MAPA e ELM, com e sem a covariável de subsídio. As metodologias foram aplicadas aos dados mensais do volume de prêmio do seguro agrícola das regiões Sul, Sudeste, Centro-Oeste e Nordeste do Brasil de 2006 a 2018. Foi observado que o modelo SARIMA univariado mostrou o melhor resultado nas regiões Centro-Oeste e Nordeste, ao passo que o SARIMA e o ELM com a covariável foram superiores na região Sul e Sudeste, respectivamente. A partir desses resultados, foi possível discutir a relevância da subvenção para a ampliação do seguro nas regiões analisadas.

**Palavras-chave:** seguro agrícola, prêmio, previsão, subvenção.

**Abstract:** The forecast of crop insurance premiums has a relevant impact on the budget dimensioning of the Crop Insurance Premium Subsidy Program (PSR). From the perspective of responsibility and public expenditure planning, it is necessary to estimate with relative precision how much is intended to be spent on the program in the short and medium-term (Constitutional Amendment no. 95, 2016). The present study makes a comparison of methods for the projection of the agricultural insurance premium, by region. The SARIMA models and the NNAR, TBATS, MAPA, and ELM algorithms were used, with and without a subsidy covariate. The methodologies were applied to the monthly data of the premium volume of crop insurance in the South, Southeast, Midwest, and Northeast regions of Brazil between 2006 and 2018. It was observed that the univariate SARIMA model showed better results in the Midwest regions and Northeast, while SARIMA and ELM with the covariate were higher in the South and Southeast, respectively. From these results, it was possible to discuss the relevance of the subsidy for the expansion of insurance in the regions analyzed.

**Keywords:** crop insurance, premium, forecast, subvention.

## 1. INTRODUÇÃO

O agronegócio brasileiro apresenta-se como um setor altamente competitivo, dinâmico e rentável. Diversos fatores conduziram o setor para esse patamar de destaque: acesso a uma gama relativamente grande de linhas de crédito rural a juros subsidiados, novas tecnologias e práticas de manejo, acesso a diferentes formas de comercialização, garantias de preços mínimos, entre outros. As soluções encontradas pelas empresas privadas são complementadas por políticas agrícolas que evoluíram ao longo de décadas.



Apesar disso, existem riscos que podem afetar a atividade agropecuária, com reflexos em toda a cadeia do agronegócio, sendo o principal deles relacionado ao risco climático (Arias et al., 2020). As adversidades climáticas podem reduzir drasticamente a produção agrícola em uma vasta extensão territorial. Nesse sentido, o Governo Federal criou o Programa de Subvenção ao Prêmio do Seguro Rural (PSR), em 2003.

Por meio do PSR, o Governo Federal subvenciona parte do prêmio do seguro agrícola, reduzindo seu preço e tornando-o mais acessível ao produtor rural. Em caso de sinistro, o seguro garante parte do valor investido no custeio da safra ou seu faturamento diante das adversidades climáticas e/ou quedas no preço do produto (Ozaki, 2008).

Uma variável-chave no PSR é o volume orçamentário necessário para sua operacionalização. A questão orçamentária é sensível por dois motivos: i) pela ótica da responsabilidade e do planejamento do gasto público, é preciso estimar com relativa precisão quanto se pretende despende com o Programa no curto e médio prazo (Emenda Constitucional nº 95, de 2016); e ii) a imprevisibilidade orçamentária causada por cortes e contingenciamentos em seus recursos. O objetivo deste estudo relaciona-se com o item i).

Complementarmente, justifica-se a importância deste estudo tendo em vista um dos itens do Relatório de Auditoria do Tribunal de Contas da União (TCU), de 2012, que trata de Auditoria Operacional no Programa de Subvenção ao PSR. Um dos itens do Relatório recomenda que o Ministério da Agricultura, Pecuária e Abastecimento (MAPA): “Desenvolva estudo de projeção da necessidade de dispêndios do PSR, para os próximos 10 anos, visando à obtenção de base para a elaboração das estimativas de aporte de recursos orçamentários nos períodos de vigência do PPA, dos planos trienais e dos orçamentos anuais, prevendo a necessidade de expansão gradual do mercado nas diversas regiões e empreendimentos agropecuários do País”.

Entre outros fatores, o equacionamento do orçamento depende da estimativa do volume total de prêmios. A importância de se determinar esse volume baseia-se no fato de que o valor da subvenção é um percentual do valor do prêmio (Mota, 2019). Isso se aplica ao orçamento do Programa como um todo. Nesse sentido, quanto maior o valor do prêmio, maiores os gastos ou os aportes de recurso público ao Programa. Os resultados obtidos no estudo de Adami & Ozaki (2012) serão atualizados e extrapolados, utilizando séries temporais mais longas e modelos estatísticos alternativos, com o objetivo de prever o valor do prêmio do seguro agrícola.

Um segundo ponto a ser observado no estudo está relacionado à dispersão espacial do PSR. Quando se observam os dados desagregados, percebe-se uma clara concentração regional. Tomando por base o período de 2014 a 2018, a região Sul foi responsável por 63% da demanda acumulada por subvenção, seguida pelas regiões Sudeste e Centro-Oeste, com 17%, Nordeste, com 2%, e Norte, com 1% (Brasil, 2021).

Notadamente, constata-se um distanciamento considerável da região Sul em relação às outras regiões. Isso provavelmente ocorre por causa de determinados fatores combinados, tais como clima adverso, frequentes quebras de safras, acultamento do seguro pelo produtor rural, forte presença das companhias seguradoras, entre outros.

Por outro lado, a demanda nas regiões Sudeste e Centro-Oeste é praticamente igual e relativamente menor em relação à região Sul, mesmo com a subvenção. Nesse contexto, ao realizar as previsões de prêmios, testar-se-á se o volume de subvenção é relevante no processo de previsão dos prêmios.

Avaliaram-se as regiões Nordeste, Centro-Oeste, Sudeste e Sul, ficando excluída a região Norte pela baixa demanda por seguros e participação na produção agrícola brasileira. Utilizou-se de métodos que trabalham com séries de tempo para a previsão do prêmio do seguro para cada uma das regiões. Além disso, o trabalho contou com uma variável mensal derivada do

volume de subvenção regional, elaborada a partir dos dados anuais de subvenção publicados pelo MAPA e utilizando a metodologia iniciada em Denton (1971) e adaptada por Sax & Steiner (2013) para transformar séries anuais em uma aproximação mensal (Brasil, 2018).

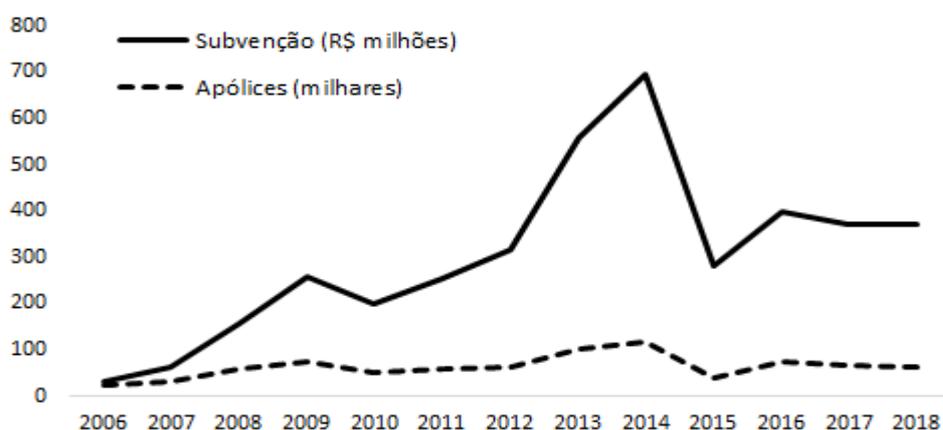
O artigo está dividido em cinco seções, além desta introdução, que é a primeira. A segunda seção faz uma breve revisão de literatura. A terceira seção traz os dados utilizados e a metodologia, detalhando os modelos e algoritmos. A quarta seção apresenta e discute os resultados. E, por fim, a última seção conclui o estudo.

## 2. REVISÃO DE LITERATURA

No Brasil, o mercado de seguro rural começou a ser desenvolvido desde a década de 1950 e 1960, com a criação da Companhia Nacional de Seguro Agrícola e do Fundo de Estabilidade do Seguro Rural (FESR). Entretanto, o seguro ganhou efetivamente mais robustez e volume a partir do final de 2005, após a criação do PSR (Ozaki, 2010).

De forma geral, a dinâmica do mercado de seguro rural operacionalizado dentro do PSR começa com os produtores contratando apólices de seguro e solicitando a subvenção federal para abater parte do prêmio cobrado pelas seguradoras. Estas, por sua vez, enviam as propostas de seguro ao MAPA, que realiza o empenho e o devido abatimento no orçamento do Programa (Guimarães & Nogueira, 2009).

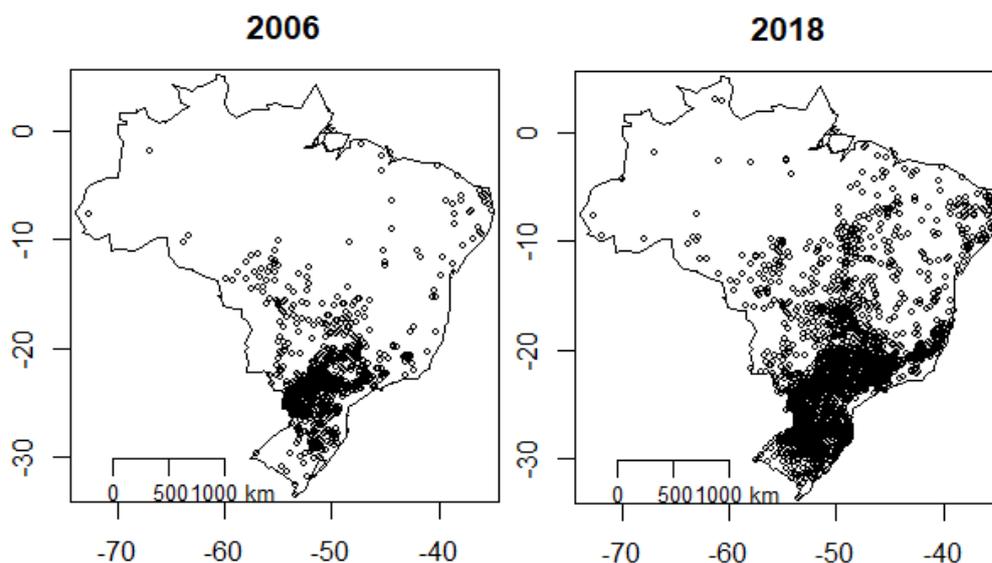
Ao longo dos anos, o número de apólices oscilou, acompanhando a evolução no volume de subvenção e exibindo uma correlação significativa<sup>1</sup> no período de 2006 a 2018 (Figura 1). Quanto maior for o volume de recurso destinado ao PSR, maior será o número de produtores contemplados pelo Programa (Medeiros, 2013; Adami & Ozaki, 2012; Santos & Silva, 2017).



**Figura 1.** Relação entre volume de subvenção anual do Prêmio do Seguro Rural (PSR), em milhões de reais a preços correntes, e número de apólices do seguro agrícola, em milhares. Fonte: elaborada pelos autores a partir dos dados do MAPA.

Ainda em relação ao número de apólices, houve uma grande evolução entre 2006 e 2018. De fato, o número de apólices passou de 21.783 para 63.554, aumentando sua presença em outros lugares, além das regiões Sudeste e Sul do país (Figura 2).

<sup>1</sup> A correlação significativa dos dados das apólices e do volume de subvenção disponíveis no MAPA foi obtida a partir do coeficiente de correlação de Pearson (1896). O coeficiente estimado foi de 0,93, rejeitando a hipótese nula de que o coeficiente é igual a zero (p-valor = 4.014e-06), com um intervalo de confiança de 0,78 a 0,98.



**Figura 2.** Distribuição de apólices do seguro agrícola no Brasil em 2006 e 2018. Fonte: elaborada pelos autores a partir dos dados do MAPA.

A literatura envolvendo a previsão de prêmios ainda é escassa. Um dos primeiros estudos foi elaborado por Adami & Ozaki (2012), que propuseram o uso de modelos estatísticos para prever o comportamento dos prêmios diretos e, com isso, calcular o montante adequado de subvenção. A partir dos resultados, os autores concluíram que o orçamento de 2011 seria insuficiente para manter a tendência de crescimento do seguro rural, reduzindo o mercado aos mesmos patamares encontrados em 2009.

Outro estudo importante utilizou o modelo ARIMA na determinação do volume ideal de subvenção para cobrir o volume de crédito rural, de forma a garantir a proteção do sistema financeiro e a massificação do seguro rural (Ozaki, 2013).

De forma mais genérica, mas no contexto do seguro agrícola, Macedo et al. (2013) mostraram a evolução do PSR no período de 2006 a 2010, ilustrando a participação regional nesse mercado e ressaltando a concentração do Programa nas regiões Sul e Sudeste do país, o que pode ser verificado também na Figura 2. Os autores ressaltam que os recursos disponibilizados para o PSR no período estudado foram insuficientes para a evolução adequada do mercado, sobretudo por contingenciamentos no MAPA em 2009 e 2010.

O presente estudo optou pela abrangência regional e pela adição de metodologias mais recentes, após os anos 2000, como aquelas que usam redes neurais para previsão. A comparação entre previsão com modelos de redes neurais e modelos SARIMA já foi feita em pesquisas envolvendo o agronegócio, embora não no seguro agrícola (Oliveira et al., 2010; Soares et al., 2015; Flores & Werner, 2019), sendo que o bom desempenho dos modelos, em termos de acurácia, reforça a hipótese de que sua utilização em outras séries que se pretendem prever, sobretudo no segmento agrícola, pode superar os resultados dos demais modelos de previsão de séries de tempo.

Cabe ressaltar também que a importância da previsão do volume de prêmios extrapola as implicações explicitadas neste estudo. A descoberta do volume futuro dos prêmios é importante

na hora de determinar o planejamento e a atuação do mercado por parte dessas seguradoras, sobretudo no planejamento de suas provisões técnicas<sup>2</sup>.

Além disso, o maior volume de informação relativa ao contexto regional contribui para a diversificação de risco das carteiras das seguradoras, principalmente considerando que o volume de prêmios é variável relevante para decisão sobre determinar o portfólio das seguradoras.

### 3. METODOLOGIA

Nesta seção, apresentam-se os dados e modelos utilizados na previsão do prêmio do seguro agrícola. São eles: i) SARIMA; ii) *Neural Network Autoregression* (NNAR); iii) *Multiple Aggregation Prediction Algorithm* (MAPA); iv) *Trigonometric Box-Cox Transformation, ARMA Errors, Trend and Seasonal Components* (TBATS); v) *Extreme Learning Machine* (ELM); e vi) modelo base (Naive).

#### 3.1 Base de Dados

Os dados utilizados compreendem as séries mensais de prêmio de seguro direto disponíveis por unidade federativa no Sistema de Estatística da SUSEP (SES). Um ponto importante a mencionar é que se considerou o volume de prêmios total do seguro agrícola, e não apenas o volume de prêmios arrecadados no âmbito do PSR (Superintendência de Seguros Privados, 2018). Adicionalmente, utilizou-se das séries de volume total de subvenção ao prêmio de seguro rural por região, disponibilizadas pelo Ministério da Agricultura, Pecuária e Abastecimento (Brasil, 2021).

Para se ter uma ideia, em 2018, o total de prêmios arrecadados pelo mercado referente ao ramo agrícola, envolvendo as 11 empresas que atuaram nesse segmento, foi de R\$ 1,05 bilhão, de acordo com a Superintendência de Seguros Privados (SUSEP), ao passo que o total de prêmios com subvenção, ou seja, que participaram do PSR, foi de R\$ 487 milhões.

Os valores foram deflacionados pelo Índice Geral de Preços - Mercado (IGP-M) do IBRE/FGV para que o modelo não tivesse que projetar a inflação dos períodos, ou seja, trabalhou-se com as séries em valores reais. A frequência dos dados era mensal e compreendeu o período entre janeiro de 2003 e dezembro de 2018, apresentando preços de janeiro de 2018.

Com o objetivo de testar as previsões do volume de prêmios pagos anualmente, retiraram-se as últimas 12 observações de cada série de tempo, e, após a seleção do modelo pela validação cruzada, projetou-se o prêmio do período retirado, que comporta os meses entre janeiro e dezembro de 2018, seguindo o procedimento utilizado em Thielbar & Dickey (2017). Todas as análises foram conduzidas com o uso do software R (R Core Team, 2013).

##### 3.1.1 Desagregando a série de subvenção

O impacto da subvenção governamental na contratação de seguro rural é conhecido na literatura (Mavroutsikos et al., 2021); portanto, é de interesse testar a relevância dessa variável na previsão dos prêmios de seguro rural nas regiões selecionadas. Dado que a frequência temporal dos dados de subvenção é anual, os valores foram desagregados temporalmente para sincronizá-los com os dados do volume de prêmios, que possuíam frequência mensal. Para isso, utilizou-se de um método de desagregação temporal para replicar um comportamento

<sup>2</sup> As provisões técnicas são um volume financeiro que as seguradoras são obrigadas a guardar como garantia de que terão saldo para cumprir suas obrigações com os sinistros.

mensal do volume anual de subvenção, ainda que não refletisse o comportamento real mês a mês da série, mas captava o nível que cada ano apresentava.

O objetivo da desagregação temporal é encontrar uma série desconhecida  $\widehat{z}_t$  de frequência  $h$ , de tal forma que sua soma resulte em uma série conhecida  $z_t$  de frequência  $l$ , em que  $l < h$ . O método apresentado em Denton (1971) utiliza uma variável  $p_t$  de frequência  $h$  e do mesmo tamanho que  $\widehat{z}_t$  como apoio para a desagregação. A configuração do método utilizado pelo trabalho é aquele descrito em Sax & Steiner (2013), em que  $p_t$  é composto apenas por  $1$ , de forma que  $\widehat{z}_t$  é definido conforme:

$$\widehat{z}_t = P_t + DU_t \quad (1)$$

em que

$$U_t = Z_t - CP_t \quad (2)$$

tal que:  $D$  é uma matriz chamada distribuição  $n \times m$ , sendo  $n$  o comprimento de  $\widehat{z}_t$  e  $m$  o de  $Z_t$ ;  $U_t$  é uma matriz  $m \times 1$ ; e  $C$  é uma matriz de conversão, conforme especificado em Sax & Steiner (2013). A determinação de  $D$  segue o proposto em Denton (1971), em que vale  $D = \Sigma C' (C \Sigma C')^{-1}$ , sendo  $\Sigma = (\Delta' \Delta)^{-1}$ , tal que  $\Delta$  é uma matriz  $n \times n$  com  $1$  na diagonal principal,  $-1$  na diagonal imediatamente abaixo da principal e  $0$  nas demais entradas.

### 3.2 SARIMA

Os modelos SARIMA (ou ARIMA sazonal) são uma extensão dos modelos ARIMA, em que uma série de tempo  $z_t$  é gerada por um processo SARIMA(p,q,d)(P,D,Q), conforme:

$$(B)\Phi(B^s)(1-B)^d(1-B^s)^D Z_t = \theta(B)\Theta(B^s)a_t \quad (3)$$

sendo que:

1.  $\phi(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - \dots - \phi_p B^p$  é o operador autorregressivo AR estacionário de ordem  $p$ ;
2.  $\theta(B) = 1 - \theta_1 B^1 - \theta_2 B^2 - \dots - \theta_q B^q$  é o operador de médias móveis MA invertível de ordem  $q$ ;
3.  $(1-B)^d$  é o processo de integração de ordem  $d$ ;
4.  $\Phi(B) = 1 - \Phi_1 B^{1s} - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}$  é o operador autorregressivo sazonal estacionário de ordem  $P$ ;
5.  $\Theta(B) = 1 - \Theta_1 B^{1s} - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs}$  é o operador de médias móveis sazonal de ordem  $Q$ ;
6.  $(1-B^s)^D$  é o operador de diferença sazonal;
7.  $a_t$  é um ruído branco;

em que:  $B$  é o operador de defasagem;  $s$  é a dimensão do período sazonal.

No caso em que há uma série de tempo  $x_t$  exógena como covariável, o modelo SARIMA é conhecido como SARIMAX (ou ARIMAX sazonal), e, como descrito em Yang et al. (1996) e Huang et al. (2005), a expressão (3) se torna:

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D Z_t = \psi(B)X_t + \theta(B)\Theta(B^s)a_t \quad (4)$$

em que:  $\psi(B)$  é o operador de defasagem e os coeficientes da covariável, valendo  $\psi(B) = 1 - \psi_1 B^1 - \psi_2 B^2 - \dots - \psi_m B^m$ .

Inicialmente, testou-se a estacionariedade das séries pelo teste de raízes unitárias Dickey-Fuller Aumentado (ADF), proposto em Dickey & Fuller (1981); depois, identificaram-se os modelos pelas chamadas função de autocorrelação (FAC) e função de autocorrelação parcial (FACP)<sup>3</sup> e selecionou-se aquele com menor AIC (*Akaike Information Criterion*)<sup>4</sup>.

Após sua estimação, fizeram-se alguns testes com os resíduos: o teste de Box & Pierce (1970) ou de Ljung & Box (1978), para verificar se não havia autocorrelação dos resíduos, e o teste de Shapiro-Wilk (1965), para verificar se seguiam uma distribuição normal.

### 3.3 NNAR

Os modelos *Neural Network Autoregression* (NNAR) são modelos híbridos de redes neurais do tipo *feedforward* e dos modelos clássicos autorregressivos (AR), constituídos de uma possível parte linear e uma não linear. A literatura costuma apresentar redes neurais como uma rede de neurônios organizados em camadas.

As entradas (*inputs*) formam a camada inferior, e as previsões (ou *outputs*), a camada superior, sendo intermediadas por uma camada oculta que contém os neurônios, os quais são valores transformados do *input* inicial por algum peso e que são transformados por funções chamadas de funções de ativação.

O componente linear é formado pelo intercepto e pela ponderação de \$\$\$ defasagens da variável objetivo, assim como nos modelos AR. A parte não linear, que contém os neurônios ocultos, transforma as variáveis imputadas, ponderadas por um parâmetro  $\gamma_{ij}$  e um viés  $\gamma_{0j}$ , por meio de uma função não linear  $\psi(\cdot)$ . Dietz (2018) e Zhang (2003) mostram que a função completa pode ser escrita como:

$$Z_t = a_0 + \sum_{i=1}^p a_i Z_{t-i} + \sum_{j=1}^n \beta_j \psi \left( \gamma_{0j} + \sum_{i=1}^p \gamma_{ij} Z_{t-i} \right) + \varepsilon_t \quad (5)$$

em que:  $\beta_j$  é a ponderação das funções não lineares; e  $\varepsilon_t$  é o componente de erro, assumido homocedástico e normalmente distribuído.

Formalmente, como mostra Hyndman (2014), é possível chamar o modelo de NNAR(p,P,k), em que:  $p$  são defasagens da série de tempo usada na entrada, como nos modelos autorregressivos;  $P$  são as defasagens sazonais; e  $k$  é o número de neurônios na camada oculta.

### 3.4 Algoritmo MAPA

O modelo conhecido como *Multiple Aggregation Prediction Algorithm* (MAPA) foi proposto por Kourentzes et al. (2014) e trabalha basicamente com a agregação da série de tempo  $Z_t$  em alguns componentes de diferentes frequências  $k$  (por exemplo, uma série mensal pode ser agregada em série bimestral, em que  $k=2$ ; trimestral, em que  $k=3$  etc.), que são projetados individualmente por métodos de suavização exponencial da família dos modelos *Exponential Smoothing* (ETS), descrita em Hyndman et al. (2008). O ETS capta os componentes estruturais de uma série de tempo, como tendências ( $T_t$ ), sazonalidades ( $S_t$ ), nível ( $L_t$ ) e ruídos ( $\varepsilon_t$ ), elaborando a projeção dos três primeiros.

<sup>3</sup> A FAC é dada por  $t, s = \text{Cov}Z_t, Z_s / \text{Var}Z_t \text{Var}Z_s$  e a FACP é dada por  $k, k$ , podendo ser encontrada ao satisfazer os sistemas das equações de Yule-Walker ( $j = k1j-1 + k2j-2 + \dots + kkj-p$ ) pela regra de Cramer.

<sup>4</sup> Trata-se de um critério para comparação e seleção de modelos, proposto em Akaike (1973).

Assim, as novas séries  $z_t^k$  são projetadas  $h$  períodos à frente por meio de seus componentes estruturais e novamente combinadas para resultar nas previsões da série original. Athanasopoulos et al. (2011) iniciaram a modelagem empírica e a avaliação do impacto no desempenho dos métodos de suavização exponencial, com a agregação de dados mensais e trimestrais para a frequência anual, realizando projeções dessas séries.

A combinação ponderada, por exemplo, a média, é comum na literatura, podendo alcançar até uma menor variância dos erros de previsão (Hibon & Evgeniou, 2005). A projeção final deverá ser aquela em que o modelo ETS minimize o critério de escolha de modelos selecionados.

### 3.5 Algoritmo TBATS

O algoritmo *Trigonometric Box-Cox Transformation, ARMA Errors, Trend and Seasonal Components* (TBATS), proposto em De Livera et al. (2011), é um avanço na proposta de Taylor (2003) para a versão linear do modelo Holt-Winters, envolvendo modificações de modelos de espaço de estados, incorporando transformações de Box-Cox, representações de Fourier com coeficientes variáveis e correção de erros do tipo ARMA, além de componentes trigonométricos que acomodam séries de tempo com sazonalidades mais complexas, bem como tratam de possíveis não linearidades:

$$Z_t^\omega = \begin{cases} \frac{Z_t^\omega - 1}{\omega}, & \text{se } \omega \neq 0 \\ \log \log Z_t, & \text{do contrário} \end{cases} \quad (6)$$

$$Z_t^\omega = L_{t-1} + \phi T_{t-1} + \sum_{i=1}^T S_{t-1}^i + d_t \quad (7)$$

$$L_t = L_{t-1} + \phi T_{t-1} + \alpha d_t \quad (8)$$

$$T_t = (1 - \phi)T + \phi T_{t-1} + \beta d_t \quad (9)$$

$$S_t^i = \sum_{j=1}^{k_i} S_{j,t}^i \quad (10)$$

$$S_{j,t}^i = S_{j,t-1}^i \cos \lambda_j^i + S_{j,t-1}^{*i} \sin \lambda_j^i + \gamma_1^i d_t \quad (11)$$

$$S_{j,t}^{*i} = -S_{j,t-1}^i \sin \lambda_j^i + S_{j,t-1}^{*i} \cos \lambda_j^i + \gamma_2^i d_t \quad (12)$$

$$d_t = \sum_{i=1}^p \phi_i d_{t-1} + \sum_{i=1}^q \theta_i \varepsilon_{t-1} + \varepsilon_t \quad (13)$$

em que:  $\omega$  é o parâmetro da transformação Box-Cox;  $L_t$  é o nível da série;  $T$  é a tendência de longo prazo;  $T_t$  é a tendência de curto prazo;  $S_j^i$  é o  $i$ -ésimo componente sazonal;  $m_1, \dots, m_t$

são os períodos sazonais;  $d_t$  é o processo  $ARMA(p,q)$ ;  $\varepsilon_t$  é um ruído branco; e  $\alpha$ ,  $\beta$  e  $\gamma_i$  são parâmetros de suavização.

Nas expressões que tratam da sazonalidade,  $\lambda_j^i = \frac{2\pi j}{m_i}$ ,  $\gamma_1^i$  e  $\gamma_2^i$  são os parâmetros de suavização,  $s_{j,t-1}^i$  são os  $i$ -ésimos componentes sazonais estocásticos e  $s_{j,t}^{*i}$  é sua denotação em termos de taxa de crescimento. Por fim, o termo  $k_i$  no somatório é dado por  $k_i = \frac{m_i}{2}$  para valores pares de  $m_i$  e por  $k_i = \frac{(m_i-1)}{2}$  para valores ímpares.

### 3.6 Algoritmo ELM

O algoritmo *Extreme Learning Machine* (ELM) foi proposto por Huang et al. (2005) e usa como base o algoritmo de redes neurais do tipo *feedforward* com camada oculta, conhecido como *Single Layer Feedforward Network*, ou SLFN (Huang & Babri, 1998).

O ELM tem uma dinâmica de camadas semelhante ao NNAR, por conter  $l$  neurônios na camada de entrada e de saída e  $i$  na camada oculta, mas não é tão dependente dos pesos e vieses da sua camada intermediária, e sim da camada final.

Assim, consideram-se a série  $Z_t$  e uma função de ativação  $G(\omega_i, b_i, Z_t)$  como a saída do  $i$ -ésimo neurônio oculto, sendo  $b_i \in R$  o chamado viés e  $\omega_i = (\omega_{i1}, \dots, \omega_{ij})$  o vetor de pesos que conecta o  $i$ -ésimo neurônio da camada oculta com o  $t$ -ésimo neurônio da camada de entrada. Portanto, pode-se considerar uma função de saída

$$\underline{z}_t = \sum_{i=1}^l \beta_i G(\omega_i, b_i, z_t) \quad (14)$$

em que:  $\beta_i = (\beta_{i1}, \dots, \beta_{ik})$  é um vetor de pesos que conecta o  $i$ -ésimo neurônio da camada oculta com o  $k$ -ésimo neurônio da camada de saída. Dessa forma, consideram-se  $Z_t$ , com  $t = 1, 2, \dots, m$ , um vetor de entrada, que serão os dados de treinamento, e  $\underline{Z}_t$  o correspondente vetor de saída, além de levar em conta  $w_i \in R^m$ , vetor de pesos, e  $b_i \in R$ , sendo ambos aleatórios. A expressão pode ser escrita na forma de matrizes, tal que  $H\beta = \underline{Z}_t$ , como mostra Singh & Balasundaram (2007), sendo necessária para a solução desse grande conjunto de equações.

O treino do algoritmo ELM, com base no SLFN, corresponde a encontrar a norma mínima da solução dos mínimos quadrados da matriz de equações  $H\beta = \underline{Z}_t$ , começando com valores aleatórios dos pesos e dos vieses na camada oculta. Calcula-se  $H^+$ , sendo essa a matriz inversa de Moore-Penrose para a matriz  $H$ , usada para calcular a solução de sistema de equações lineares que não possuem uma solução única, ou então a norma mínima de tal sistema. Depois, encontra-se a matriz de pesos  $\underline{\beta}$ , dada por  $\underline{\beta} = H^+ \underline{Z}_t$ , e, por fim, para cada entrada  $z_t$ , a saída

$$\underline{z}_t \text{ é dada pela função } \underline{z}_t = \sum_{i=1}^l \beta_i G(\omega_i z_t + b_i).$$

### 3.7 Naive (modelo base)

Seguindo Pappenberger et al. (2015), optou-se por comparar os métodos descritos anteriormente com um modelo mais básico possível, de fácil reprodução e sem complexidade computacional, que serve de base de comparação (*baseline*).

O objetivo desse procedimento é que, além de escolher o algoritmo que produz o menor erro de predição entre os métodos apresentados, a abordagem selecionada deve também superar o desempenho de um algoritmo simples. O *Naive* sazonal com deslocamento descrito

em Hyndman (2014) foi o método considerado como *baseline*. A projeção de  $h$  períodos à frente de  $z_t$  é dada por

$$\underline{Z}_{t+h} = Z_{t+h-m(k+1)} + \frac{h}{t-1} \sum_{i=2}^n (Z_i - Z_{i-1}) = Z_{t+h-m(k+1)} + h \left( \frac{Z_t - Z_1}{t-1} \right) \quad (15)$$

em que:  $m$  é o período sazonal (12, no caso de séries mensais); e  $k$  é o inteiro resultante de  $\frac{h-1}{m}$ . Em suma, o algoritmo assume que a projeção  $h$  períodos à frente é determinada por um valor passado já conhecido (a primeira parte do termo do lado direito da expressão anterior), acrescido de uma constante de deslocamento.

### 3.8 Critério de seleção de modelos

O critério de seleção de modelos escolhido foi a Raiz do Erro Quadrático Médio (RMSE, na sigla em inglês) como método de acurácia (Hyndman & Koehler, 2006). Dadas uma série  $z_t$  e sua previsão  $\underline{z}_t$ , pode-se definir a métrica conforme a expressão:

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (Z_t - \underline{z}_t)^2} \quad (16)$$

Desta forma, quanto menor for tal métrica, melhor será o valor do ajuste da previsão em relação à série original, servindo como um parâmetro para tomar a decisão entre modelos de metodologias distintas (Alexander, 2007).

### 3.9 Validação cruzada (VC) e índice de habilidade (Skill)

O cálculo do RMSE foi avaliado no contexto de um processo conhecido como validação cruzada, que permite a confirmação da capacidade de generalização do algoritmo utilizado (Bergmeir et al., 2018). A validação cruzada para séries de tempo proposta em Hyndman (2014) diferencia-se do processo de validação cruzada utilizado em outros tipos de conjuntos de dados, como aqueles em corte transversal.

Na abordagem proposta, o conjunto de dados de treino é dividido em subconjuntos diferentes, em que existe um primeiro conjunto com um número mínimo  $k$  de observações da série, e os próximos conjuntos dispõem de uma observação a mais que o anterior. O processo consiste em selecionar  $k+i$  observações da série para treino e, desse conjunto, utilizar  $k+i+h$  para o teste do modelo, elaborando a projeção das últimas  $h$  observações e computando o erro de projeção.

Ao todo, se o conjunto tiver  $n$  observações, então a validação cruzada terá  $n-h-k$  cortes de tempos testados e um valor de RMSE para cada teste, que são compilados em uma média final. Para facilitar o confronto dos métodos para cada região com o baseline, adotou-se o índice de proficiência (ou *Skill*), proposto em Pappenberger et al. (2015), que vale:

$$Skill = \frac{f(\underline{Z}_t^i)}{f(\underline{Z}_t^b)} \quad (17)$$

em que:  $\underline{z}_t^i$  são as projeções do algoritmo  $i$ ;  $\underline{z}_t^b$  são as projeções do *baseline*; e  $f(\cdot)$  é uma métrica de performance, que, no presente trabalho, é representada pelo RMSE. Quanto menor o valor do RMSE e do *Skill*, melhor a capacidade preditiva do modelo. Especificamente em relação ao *Skill*, caso o valor da métrica seja igual a 1, o modelo tem capacidade preditiva igual ao do modelo *baseline*; caso seja inferior a 1, a capacidade preditiva é superior; e caso seja superior a 1, a capacidade preditiva é inferior.

#### 4. Resultados e Discussão

Iniciando pela avaliação dos modelos para cada região no caso univariado, a Tabela 1 apresenta o resultado daqueles que exibiram a melhor medida de acurácia, compreendendo o RMSE das projeções dos prêmios de seguro agrícola do período de janeiro de 2006 a dezembro de 2017 pela validação cruzada, além das estimativas do *Skill* a partir do RMSE.

No caso dos modelos SARIMA estimados, todos os coeficientes dos componentes AR, MA e MA sazonal mostraram-se estatisticamente significantes. Já os testes Ljung-Box e Box-Pierce não rejeitaram a hipótese nula de erros independentes e identicamente distribuídos. Por fim, o teste Shapiro-Wilk rejeitou a hipótese de não normalidade dos resíduos, e o AIC foi o menor possível.

**Tabela 1.** Resultados das medidas de acurácia das projeções por região, por modelo univariado, calculados a partir da validação cruzada

	Sul		Sudeste		Centro-Oeste		Nordeste	
	RMSE	Skill	RMSE	Skill	RMSE	Skill	RMSE	Skill
SARIMA	19,19	0,93	10,76	0,80	15,10	0,93	1,65	0,47
NNAR	25,76	1,25	14,71	1,09	16,83	1,03	3,53	1,01
ELM	17,79	0,86	12,58	0,93	15,17	0,93	3,34	0,95
TBATS	35,21	1,71	15,35	1,14	23,95	1,47	6,66	1,90
MAPA	24,91	1,21	17,61	1,30	21,24	1,30	4,67	1,33
Naive	20,58	1,00	13,50	1,00	16,31	1,00	3,51	1,00

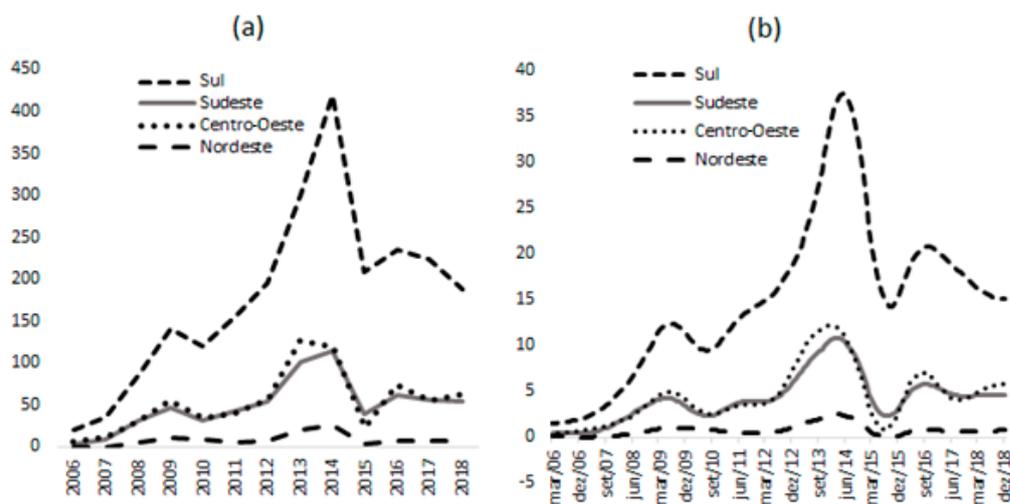
Fonte: elaborada pelos autores.

O destaque é que nenhuma configuração dos algoritmos TBATS e MAPA superaram o desempenho do *baseline* em termos de RMSE para as regiões, refutando a hipótese de que esses métodos possam contribuir para a projeção das séries de prêmios. A média dos RMSE resultantes do processo de validação cruzada desses algoritmos equivale a pelo menos 115% do resultado do *Naive* sazonal, no caso do Sudeste, chegando até 190%, no caso do Nordeste.

Considerando a região Sul, o modelo SARIMA mostrou bom resultado, superando o *baseline*, mas ficando abaixo do algoritmo ELM, que apresentou um RMSE equivalente a 86% daquele exibido pelo *baseline*. No caso do Sudeste, o SARIMA e o ELM também superaram o *baseline*, mas o primeiro método mostrou a melhor acurácia em suas projeções, registrando apenas 80% do RMSE exibido pelo *Naive* sazonal.

Resultado similar foi encontrado nos casos do Centro-Oeste e Nordeste, com bom desempenho do SARIMA e do ELM, mas com o primeiro mostrando melhor performance em relação aos demais. A diferença de nível de RMSE entre as regiões ocorre pela diferença de nível das próprias séries, sendo que a região Sul apresentou o maior volume de prêmio ao longo do ano, e o Nordeste, o menor.

Por sua vez, no caso em que os métodos foram estimados com a covariável de subvenção, foi necessário estimar a variável derivada da subvenção mensal por região pelo método descrito em 2.1 (Figura 3).



**Figura 3.** (a) Série de subvenção anual do seguro agrícola, por região, em milhões de reais, deflacionado pelo IGP-M, com ano-base em 2018; (b) Série mensal derivada da subvenção do seguro agrícola anual, por região, em milhões de reais, deflacionado pelo IGP-M, com mês-base em janeiro de 2018, utilizando o método proposto por Sax & Steiner (2013). Fonte: elaborada pelos autores.

Os algoritmos TBATS e MAPA não permitem a projeção com covariável, mas, como não apresentaram baixo RMSE, mesmo no caso univariado, não houve prejuízo para a avaliação ao não considerá-los. A Tabela 2 exhibe os resultados da validação cruzada dos métodos no caso em que a variável estimada foi utilizada, sendo que, dessa vez, tanto o ELM quanto o SARIMA não superaram o *baseline* (que também foi estimado com a covariável) em todas as regiões.

Na região Sul, diferentemente do caso univariado, o SARIMA com covariável mostrou os menores RMSE e *Skill*, ao passo que o melhor método, no caso do Sudeste, mudou para o ELM, com o *Skill* de ambos próximo a 0,70. O Nordeste também teve o SARIMA como melhor método, mas os seus resultados não superaram aqueles do caso univariado, embora com um resultado muito próximo. Apenas o Centro-Oeste mostrou o NNAR como melhor método de predição durante a validação cruzada, no caso em que foi utilizada a covariável de subvenção, com 15,12 de RMSE e superando sutilmente o caso univariado (15,17 de RMSE), embora o *Skill* tenha sido menor, visto que o *baseline* mostrou melhor desempenho quando considerada a covariável.

**Tabela 2.** Resultados das medidas de acurácia das projeções por região e por modelo com covariável mensal estimada para o subsídio

	Sul		Sudeste		Centro-Oeste		Nordeste	
	RMSE	Skill	RMSE	Skill	RMSE	Skill	RMSE	Skill
SARIMA	13,90	0,70	9,90	0,76	19,13	1,22	1,74	0,52
NNAR	24,57	1,25	13,12	1,00	15,12	0,97	3,32	0,99
ELM	15,69	0,80	9,79	0,75	15,29	0,98	3,34	1,00
Naive	19,73	1,00	13,08	1,00	15,65	1,00	3,36	1,00

Fonte: elaborada pelos autores.

Desta forma, podem-se selecionar os quatro métodos que mostraram melhor performance preditiva, considerando os casos univariados e com covariáveis: no caso da região Sul, destacaram-se o SARIMA(0,0,0)(1,1,0) e a covariável com o menor RMSE; na região Sudeste, o ELM com covariável; no Centro-Oeste, o SARIMA(1,0,0)(1,1,2)<sup>5</sup> univariado, com menor RMSE; e, por fim, no Nordeste, o SARIMA(13,0,0)(2,1,0)<sup>6</sup> univariado – ambos os modelos sem a covariável significativa. Os resultados dos coeficientes e testes estão dispostos na Tabela 3.

**Tabela 3.** Estimação dos parâmetros dos modelos SARIMA para o volume de prêmio do seguro agrícola mensal por região do Brasil entre 2006 e 2018 (erro-padrão)

	Sul	Nordeste	Centro-Oeste
	SARIMA (0,0,0)(1,1,0)	SARIMA (13,0,0)(2,1,0)	SARIMA (1,0,0)(1,1,2)
Subvenção	0,576*** (0,201)	-	-
AR(1)	-	0,472*** (0,078)	0,496*** (0,077)
AR(11)	-	0,177** (0,090)	-
AR(13)	-	-0,075* (0,076)	-
SAR(1)	-	-0,612*** (0,095)	-0,567*** (0,098)
SAR(2)	-0,179** (0,087)	-0,321*** (0,091)	-
SMA(2)	-	-	-0,339*** (0,117)
Observações	132	132	132
Teste de Box-Pierce (autocorrelação dos resíduos)	0,1817	0,8662	0,1587
Teste Ljung-Box (lag 1,3,6,9)	0,14; 0,11; 0,16; 0,22	0,75; 0,54; 0,70; 0,79	0,54; 0,54; 0,45; 0,14
Teste Shapiro-Wilk (Normalidade)	0,0001275	1,036 x 10 <sup>-9</sup>	2,658 x 10 <sup>-5</sup>

**Nota:** (\*) denota a rejeição de H0 a 10% de significância; (\*\*) denota a rejeição de H0 a 5% de significância; e (\*\*\*) denota a rejeição de H0 a 1% de significância. Fonte: elaborada pelos autores a partir dos dados da pesquisa

#### 4.1 Utilização prática dos métodos escolhidos

Assim, cabe ressaltar que o resultado encontrado vai em direção ao registrado em Adami & Ozaki (2012), que apontam o modelo SARIMA como o mais indicado para a predição da série de volume de prêmios totais, sendo o caso para três das quatro séries abordadas. A diferença no presente trabalho é que dois métodos selecionados contam com a covariável estimada, ou seja, com a presença da subvenção. Os resultados sugerem, portanto, que a subvenção tem impacto nas regiões Sul e Sudeste, mas não foi significativa nas regiões Centro-Oeste e Nordeste pelo critério RMSE.

<sup>5</sup> Considerando apenas as defasagens 1 do AR, 1 do AR sazonal e 2 do MA sazonal.

<sup>6</sup> Considerando apenas as defasagens 1, 11 e 13 do AR, além do 1 e 2 do AR sazonal.

Nota-se também que o método ELM acompanha de forma muito próxima a performance do SARIMA, sendo que apenas no caso do Nordeste o RMSE foi notadamente superior. Após a seleção, estimaram-se as previsões dos volumes de prêmio entre janeiro de 2018 e dezembro de 2018 com o melhor método de cada região (Tabela 4).

**Tabela 4.** Comparação entre valor original e predito das séries de prêmios por região do país para os meses de 2018, em milhões de reais, de acordo com cada modelo selecionado

	Sul		Sudeste		Centro-Oeste		Nordeste	
	Projeção	Original	Projeção	Original	Projeção	Original	Projeção	Original
jan./18	52,3	37,0	15,4	24,0	43,6	51,9	2,3	5,3
fev./18	36,2	32,2	15,1	13,6	17,1	19,2	3,2	2,9
mar./18	70,4	70,5	32,2	32,9	43,3	43,4	4,9	3,9
abr./18	58,2	88,1	18,7	37,9	35,1	46,4	4,6	6,9
maio/18	85,3	84,7	33,2	43,5	33,1	42,5	6,6	7,8
jun./18	110,1	112,7	42,1	35,6	38,1	51,3	9,7	14,2
jul./18	94,0	114,0	38,3	37,9	30,2	38,3	14,0	13,2
ago./18	107,8	118,6	47,7	48,9	35,0	46,0	15,1	15,2
set./18	81,8	72,9	43,6	55,4	28,0	51,9	13,9	13,2
out./18	74,6	54,7	56,3	62,5	43,4	50,4	11,1	10,5
nov./18	49,1	32,1	41,8	18,1	38,6	48,8	8,0	7,6
dez./18	11,0	4,8	17,6	20,0	16,5	14,3	6,0	6,1

**Fonte:** elaborada pelos autores.

A previsão dos meses de 2018 da região Sul resultou em um RMSE de 14,34, um pouco superior ao resultado durante a validação cruzada (13,90), com subestimação em abril, julho e agosto. Por sua vez, o volume de prêmio projetado para a região Sudeste resultou em um RMSE de 10,56, também superando um pouco aquele visto durante a validação cruzada, de 9,79, subestimando os resultados efetivos verificados em abril, maio e setembro.

Com RMSE de 10,72 e validação cruzada de 15,10, a projeção do prêmio para a região Centro-Oeste foi subestimada em boa parte do ano de 2018, acertando o movimento, mas errando o nível. Por fim, a previsão da região Nordeste para esse período gerou um RMSE de 1,79, muito próximo do 1,65 obtido durante a validação cruzada, com maior desvio verificado em junho.

Com os métodos apresentados, as seguradoras podem estimar o volume de prêmio esperado para cada região do país e, ao conhecerem uma estimativa mais robusta da demanda regional, elaborar estratégias de negócios mais assertivas. Pelo lado do governo, os métodos aqui empregados podem auxiliar no cálculo do dimensionamento do orçamento necessário para operacionalizar o PSR.

## 5. Conclusões

O mercado de seguro rural, sobretudo na modalidade agrícola, tem se expandido desde 2005, com forte apoio do Governo Federal. Ainda assim, o ritmo de crescimento e a abrangência do Programa são pequenos quando comparados à dimensão do agronegócio brasileiro. Um dos principais problemas do PSR é a imprevisibilidade orçamentária causada por cortes e contingenciamentos em seus recursos.

Em qualquer política pública, no âmbito do Governo Federal, é fundamental saber qual será o universo de indivíduos a serem beneficiados e quanto será gasto em recurso público até se atingir a meta estabelecida. No caso do PSR, tanto o número total de beneficiados quanto o gasto público total não estão claros.

Uma forma de prever o total de recursos gastos no Programa é estimando-se o prêmio total do mercado de seguro agrícola, com e sem subvenção. Nesse contexto, o objetivo do estudo foi propor modelos que forneçam boas previsões do prêmio em nível regional. Os resultados corroboram aqueles encontrados em Adami & Ozaki (2012).

O modelo SARIMA, com e sem covariáveis, mostrou-se o mais eficaz na elaboração das previsões dos prêmios. Ademais, testou-se a hipótese da relevância da subvenção na contratação do seguro. Para isso, derivaram-se variáveis mensais do volume de subvenção disponibilizado pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA), utilizando-se do método proposto em Sax & Steiner (2013), a partir da variável de frequência anual. No entanto, apenas em dois casos essas variáveis se mostraram eficazes na redução do erro de predição.

Selecionado a partir do processo de validação cruzada, o modelo SARIMA(0,0,0)(1,1,0) foi o menor RMSE entre os métodos testados para a região Sul, além de utilizar a variável estimada para a subvenção mensal. A região Sudeste teve o *Extreme Learning Machine* (ELM) como o melhor método de predição, também com a sua respectiva variável de subvenção. Por sua vez, para os casos do Centro-Oeste e do Nordeste, o modelo univariado mostrou a melhor acurácia, com as configurações SARIMA(1,0,0)(1,1,2) e SARIMA(13,0,0)(2,1,0) como as melhores opções, respectivamente.

Esses resultados sugerem que a subvenção nas regiões Centro-Oeste e Nordeste tem pouca influência nos modelos de previsão. De fato, quando se observa a penetração do PSR nas regiões, os recursos da subvenção são direcionados, em sua grande maioria, para o Sul do país. Por outro lado, as regiões Sudeste e Centro-Oeste são responsáveis por apenas 17% cada uma.

No entanto, a região Centro-Oeste possui área plantada média com as principais culturas seguradas igual a 22,3 milhões de hectares, ao passo que a região Sudeste possui área igual a 4,5 milhões de hectares, ou seja, um valor consideravelmente menor (Instituto Brasileiro de Geografia e Estatística, 2021). Isso mostra a baixa penetração do seguro subvencionado no Centro-Oeste, como apontam os resultados.

Tais resultados, porém, merecem cuidado ao serem interpretados. Ao observar-se o valor acumulado da subvenção por estado, entre 2014 e 2018, nota-se que dois estados do Centro-Oeste, Mato Grosso do Sul e Goiás, estão entre os 10 maiores, respectivamente, em quinto e sexto lugares com subvenções iguais a R\$ 156,2 milhões e R\$ 113,9 milhões. Supostamente nesses dois estados a subvenção tem relativo impacto na contratação do seguro. Portanto, a escala espacial utilizada no estudo impossibilita a análise das peculiaridades em um nível mais desagregado, o que é uma limitação deste trabalho.

Outros fatores não abordados neste estudo também podem influenciar a demanda por seguro agrícola. Em escala local, a presença de corretores especializados, por exemplo, é fundamental não apenas para massificar as vendas, mas também fornecer valiosas informações sobre o funcionamento do seguro e seus benefícios. A correta mensuração dos prejuízos e a intempestividade no pagamento das indenizações são essenciais para demonstrar a importância do seguro e justificar a subsequente contratação. Não obstante, em diversos locais, o clima pode ser estável com um bom regime de chuvas ao longo da safra. Nesse caso, outros tipos de riscos serão mais importantes. Como consequência, o seguro não será atrativo aos produtores.

Por fim, os modelos e os resultados deste estudo podem ser úteis ao mercado no momento de planejar ações de ampliação da carteira de forma mais assertiva. Para o Governo Federal, obter uma estimativa da previsão de prêmios pode nortear os gestores no dimensionamento do orçamento do PSR no médio e longo prazo. Portanto, é importante ampliar os estudos com foco na desagregação espacial dos dados e no teste de modelos de previsão alternativos àqueles utilizados neste estudo.

## Referências

- Adami, A., & Ozaki, V. (2012). Modelagem estatística dos prêmios do seguro rural. *Revista de Política Agrícola*, 21(1), 60-75.
- Akaike, H. (1973). Information theory and the maximum likelihood principle. In *2nd International Symposium on Information Theory*. Budapest.
- Alexander, G. D. (2007). *Simple kNN-method for times series prediction* (p. 2). Hamburg: Neural Forecasting Competition.
- Arias, D., Mendes, P., & Abel, P. (2020). *Revisão rápida e integrada da gestão de riscos agropecuários no Brasil*. Recuperado em 22 de fevereiro de 2021, de <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/142661/1/Revisao-rapida-e-integrada-da-gestao.pdf>
- Athanasopoulos, G., Hyndman, R. J., & Song, H. (2011). The tourism forecasting competition. *International Journal of Forecasting*, 27(3), 822-844.
- Bergmeir, C., Hyndman, R., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120, 70-83.
- Box, G., & Pierce, D. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65, 1509-1526.
- Brasil. Ministério da Agricultura, Pecuária e Abastecimento. (2018). *Relatório seguro rural: resultado geral 2018*. Recuperado em 22 de fevereiro de 2021, de <http://www.agricultura.gov.br/assuntos/riscosseguro/seguro-rural/relatorios-estatisticos>
- Brasil. Ministério da Agricultura, Pecuária e Abastecimento. (2021). *Atlas do seguro rural*. Recuperado em 22 de fevereiro de 2021, de <http://indicadores.agricultura.gov.br/atlasdoseguro/index.htm>
- Brasil. Instituto Brasileiro de geografia e Estatística. (2021). Sistema IBGE de Recuperação Automática - SIDRA, v. 2021.
- De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496), 1513-1527.
- Denton, F. (1971). Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization. *Journal of the American Statistical Association*, 66(333), 99-102.
- Dickey, D., & Fuller, W. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49(4), 1057-1072.
- Dietz, D. (2018). *Autoregressive neural network processes: univariate, multivariate and cointegrated models with application to the German automobile industry* (Tese de doutorado). Recuperado em 22 de fevereiro de 2021, de <https://opus4.kobv.de/opus4-uni-passau/frontdoor/index/index/docId/142>
- Flores, J. H., & Werner, L. (2019). *Aplicação de redes neurais artificiais à previsão de vendas de máquinas agrícolas: um estudo de caso junto a AGCO do Brasil Ltda* (Trabalho de conclusão de curso). Recuperado em 22 de fevereiro de 2021, de <https://lume.ufrgs.br/handle/10183/135441>
- Guimarães, M. F., & Nogueira, J. M. (2009). A experiência norte-americana com o seguro agrícola: lições ao Brasil? *Revista de Economia e Sociologia Rural*, 47(1), 27-58.
- Hibon, M., & Evgeniou, T. (2005). To combine or not to combine: selecting among forecasts and their combinations. *International Journal of Forecasting*, 21(1), 15-24.

- Huang, C., Huang, C., & Wang, M. A. (2005). Particle Swarm Optimization to Identifying the ARMAX Model for Short-Term Load Forecasting. *IEEE Transactions on Power Systems*, 20(2), 1126-1133.
- Huang, G., & Babri, H. A. (1998). Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions. *IEEE Transactions on Neural Networks*, 9, 224-229.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679-688.
- Hyndman, R. J., Koehler, A. B., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Springer.
- Hyndman, R. J. (2014). Measuring forecast accuracy. In: J. Wiley & S. Hoboken. Business forecasting: Practical problems and solutions, p. 177-183.
- Kourentzes, N., Petropoulos, F., & Trapero, J. R. (2014). Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, 30(2), 291-302.
- Ljung, G., & Box, G. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297-303.
- Macedo, L. O., Pacheco, A. B., & Santo, E. (2013). A evolução do Programa de Subvenção do Prêmio do Seguro Rural: uma avaliação do período 2006-10. *Revistas Eletrônicas FEE*, 40(4), 99-112.
- Mavroutsikos, C., Giannakas, K., & Walters, C. (2021). The role of premium subsidies in crop insurance. *PLoS One*, 16(4), e0250129.
- Medeiros, E. A. (2013). Avaliação da implementação do programa de subvenção do prêmio do Seguro Rural. *Revista de Economia e Sociologia Rural*, 51(2), 1-14.
- Mota, A. A. L. (2019). *Previsão de prêmio e a ocorrência de sinistros no mercado de seguro agrícola brasileiro* (Dissertação de mestrado). Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba. <http://dx.doi.org/10.11606/D.11.2019.tde-19112019-181327>.
- Oliveira, A. C., Souza, A. A., Lacerda, W. S., & Gonçalves, L. R. (2010). Aplicação de redes neurais artificiais na previsão da produção de álcool. *Revista Ciência e Agrotecnologia*, 34(2), 279-284.
- Ozaki, V. A. (2008). Em busca de um novo paradigma para o seguro rural no Brasil. *Revista de Economia e Sociologia Rural*, 46(1), 1-23.
- Ozaki, V. A. (2010). Uma digressão sobre o programa de subvenção ao prêmio do seguro rural e as implicações para o futuro do mercado do seguro rural. *Revista de Economia e Sociologia Rural*, 48(4), 757-776.
- Ozaki, V. A. (2013). Qual o custo governamental para segurar todo o crédito de custeio das lavouras? *Revista de Economia e Sociologia Rural*, 51(1), 125-138.
- Pappenberger, F., Ramos, M. H., Cloke, H. L., Wetterhall, F., Alfieri, L., Bogner, K., Mueller, A., & Salamon, P. (2015). How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction. *Journal of Hydrology*, 522, 697-713.
- R Core Team. (2013). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Santos, G., & Silva, F. (2017). *Dez anos do Programa de Subvenção ao Prêmio de Seguro Agrícola: proposta de índice técnico para análise do gasto público e ampliação do seguro* (Texto para Discussão, No. 2290). IPEA.

- Sax, C., & Steiner, P. (2013). Temporal disaggregation of time series. *The R Journal*, 5(2).
- Singh, R. & Balasundaram, S. (2007). Application of extreme learning machine method for time series analysis. *International Journal of Intelligent Technology*, 2(4), 256-262.
- Soares, F. C., Robaiana, A. D., Peiter, M. X., & Russi, J. L. (2015). Predição da produtividade da cultura do milho utilizando rede neural artificial. *Ciência Rural*, 45(11), 1987-1993.
- Superintendência de Seguros Privados – Susep. (2018). *Dados de prêmio do seguro agrícola por Unidade Federativa*. Recuperado em 22 de fevereiro de 2021, de <http://www2.susep.gov.br/menuestatistica/SES/premiosdesegurosporuf.aspx?id = 4>
- Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, 54(8), 799-805.
- Thielbar, M., & Dickey, D. (2017). Neural networks for time series forecasting. *Future Computing and Informatics Journal*, 2(1), 39-47.
- Yang, H., Huang, C., & Huang, C. (1996). Identification of ARMAX model for short term load forecasting: an evolutionary programming approach. *IEEE Transactions on Power Systems*, 11(1), 403-408.
- Zhang, P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159-175.

**Recebido:** Fevereiro 22, 2021.

**Aceito:** Junho 24, 2021.

**JEL Classification:** G22, Q13, Q18.