



Congresso Internacional de Administração
ADM 2020

Administração Ágil
Inovação e Trabalho Remoto

25 a 27
de outubro

Ponta Grossa - Paraná - Brasil

PREVISÃO E PLANEJAMENTO: UM MODELO DE PROJEÇÃO PARA ARRECADAÇÃO DO ESTADO DO PARANÁ

FORECAST AND PLANNING: A PROJECTION MODEL FOR COLLECTION IN THE STATE OF PARANÁ.

ÁREA TEMÁTICA: ADMINISTRAÇÃO PÚBLICA

Mario Filizzola-Costa, UEM, Brasil, mariofilizzola@gmail.com

Marcos Junio Ferreira de Jesus, UNESPAR, Brasil, marcos_junio@hotmail.com

Eloisa Paula de Oliveira, UNESPAR, Brasil, eloisapauladeoliveira@gmail.com

Rodrigo Monteiro da Silva, UEM, Brasil, rodrygomsylva@gmail.com

Resumo

A aprovação da Lei de Responsabilidade Fiscal – LRF em 2000, associado à crise financeira de 2008, que afetou sobremaneira a gestão fiscal, tem tornado a utilização de metodologias estatísticas de previsão para indicadores fiscais, de extrema importância a gestão pública. Desta forma, o presente estudo tem como objetivo principal desenvolver um modelo de previsão para a arrecadação do estado do Paraná, baseado no método SARIMAX, uma variação da metodologia de Box e Jenkins (1976) para séries sazonais SARIMA, que considera a inclusão de variáveis exógenas ao modelo, além de realizar um comparativo entre os dois métodos indicados. As variáveis exógenas incluídas ao modelo SARIMAX são: o índice de atividade econômica do Banco Central para o estado do Paraná, e o índice de inflação IPCA. Os resultados indicaram uma maior eficiência do modelo SARIMAX em relação ao modelo SARIMA, além de demonstrar uma boa acuracidade e eficiência preditiva para 12 meses. Em relação as variáveis exógenas, as mesmas apresentaram uma relação diretamente proporcional frente a série de arrecadação do estado.

Palavras-chave: Lei de Responsabilidade Fiscal; Gestão Pública; Modelo de Previsão; Modelo SARIMA; Modelo SARIMAX.

Abstract

The approval of the Fiscal Responsibility Law – LRF in 2000, associated with the 2008 financial crisis, which greatly affected fiscal management, has made the use of statistical forecasting methodologies for fiscal indicators extremely important to public management. Thus, the main objective of the present study is to develop a forecast model for the state of Paraná's tax revenue, based on the SARIMAX method, a variation of the methodology of Box and Jenkins (1976) for seasonal series SARIMA, which considers the inclusion of variables exogenous to the model, in addition to making a comparison between the two indicated methods. The exogenous variables included in the SARIMAX model are: the Central Bank's economic activity index for the state of Paraná, and the IPCA inflation index. The results indicated a greater efficiency of the SARIMAX model compared to the SARIMA model, in addition to demonstrating good accuracy and predictive efficiency for 12 months. Regarding the exogenous variables, they showed a directly proportional relationship with the state's tax collection series.

Keywords: Fiscal Responsibility Law, Public Management, Forecast Model, SARIMA Model, SARIMAX Model.

1. INTRODUÇÃO

A implantação da Lei de Responsabilidade Fiscal (LRF) em 2000, dotando a administração pública de novas responsabilidades, tornou a adoção de metodologias de previsão imprescindíveis para o processo de gestão, dada a diminuição do risco inerente às incertezas

macroeconômicas. Já após a crise de 2008, o interesse por análises de previsão para indicadores econômicos fiscais potencializou-se, devido às instabilidades não previstas nos orçamentos dos países e estados.

As metodologias existentes comumente utilizadas são: os modelos de Box e Jenkins (1976) e suas extensões, como o ARIMA e o SARIMA; os modelos de vetores autorregressivos e suas extensões, como o VAR e o VEC; e os modelos de redes neurais, como o NNARX. Mas, uma metodologia tem tomado espaço na literatura internacional, os modelos ARIMAX e SARIMAX, considerados uma extensão dos modelos padrões de Box e Jenkins que consideram o acréscimo de variáveis exógenas ao modelo univariado.

Considerando a escassez de estudos sobre o tema na literatura nacional, ainda mais se utilizando desta metodologia, o presente artigo busca desenvolver um modelo de previsão para a arrecadação do estado do Paraná, comparando o tradicional modelo de Box e Jenkins com sazonalidade SARIMA ao modelo SARIMAX. As variáveis exógenas utilizadas no modelo proposto se atém ao índice de atividade econômica do Banco Central para o Estado do Paraná, o IBC-Br, e o índice de inflação do IBGE, o IPCA.

As hipóteses levantadas frente ao resultado do modelo estatístico, considera que tanto o índice de atividade quanto o índice de inflação devem apresentar um comportamento diretamente proporcionais em relação aos níveis de arrecadação do estado, demonstrando assim que quanto maior for a atividade econômica e a inflação maior será a arrecadação e vice-versa.

Desta forma, o artigo está dividido em 4 seções além desta introdução, com a primeira seção apresentando o referencial teórico, a segunda seção demonstrando os procedimentos metodológicos adotados, na terceira seção a análise de dados e na quarta seção a conclusão.

2. PREVISÃO DA ARRECADAÇÃO DE RECEITAS FEDERAIS

No cenário macroeconômico da arrecadação tributária, no Brasil, uma previsão bem feita reverte-se de fundamental importância (Campos, 2009). Vale registrar, que a aprovação da Lei de Responsabilidade Fiscal (LRF) em 2000 pelo Congresso Nacional, dotou o administrador público de novas responsabilidades com relação aos orçamentos da União, dos estados e municípios, devido às restrições/limitações orçamentárias (Campos, 2009; Aguilar, 2012).

Nessa nova cultura de disciplina fiscal para os três poderes (executivo, legislativo e judiciário) “criou rigorosos dispositivos para acompanhar a obediência à lei mencionada e severas punições pelo seu descumprimento [...] para os gestores públicos e ordenadores de despesa” (Aguilar, 2012, p. 233).

No âmbito da administração pública, o grande desafio é equilibrar as contas públicas. Para isso, segundo Almeida e Ferreira (2018, p. 23) “há a obrigatoriedade de se proceder a previsão das receitas, a qual, combinada com a redução dos índices de inflação permitiu que o orçamento deixasse de ser uma peça de ficção, transformando-se um elemento estratégico e de controle financeiro”.

Neste contexto, uma das alternativas que vem sendo adotada pelo governo federal como apoio à tomada de decisões dos gestores é a utilização de metodologias de séries temporais para a previsão da arrecadação de receitas federais (Campos, 2009).

Poucos trabalhos têm tratado sobre o tema, no Brasil. A recente disponibilidade de grandes conjuntos de dados econômicos para a realização de previsões de variáveis econômicas, requer a aplicação de técnicas capazes de utilizar essas informações de forma eficiente (Barbosa, Ferreira, & Silva, 2020).

Ferreira, Bierens e Castelar (2005) destacam o uso de modelos fatoriais lineares e não lineares para prever a taxa de crescimento do PIB brasileiro e reportam que esses modelos geram previsões com menores erros quadráticos médios de previsão (EQMP) do que as dos modelos de vetores autorregressivos (VAR) e auto-regressivos (AR).

Kejllin (2020, p. 36) explica que os modelos para séries temporais podem ser classificados em modelos paramétricos, e modelos não-paramétricos. Nos modelos paramétricos o número de parâmetros é finito, visto que “a análise é feita no domínio do tempo”. Já, os modelos não-paramétricos envolvem um número infinito de parâmetros.

Dentre estes modelos, os mais frequentemente utilizados são os modelos auto-regressivo (AR), o auto-regressivo e de média móvel (ARMA), o modelo auto-regressivo integrado e de média móvel (ARIMA), modelos de memória longa (ARFIMA), modelos estruturais e modelos não-lineares (Kejllin, 2020, p. 36). Os modelos não-paramétricos mais usados são a função de auto-covariância e sua transformada de Fourier (Morettin & Toloi, 2006).

É importante notar, que a previsão da arrecadação de receitas federais está implícita na Lei de Responsabilidade Fiscal (LRF), sendo imprescindível a cada um dos entes federativos desenvolver métodos confiáveis de previsão de suas receitas e despesas. A responsabilidade na gestão fiscal deve ser prioridade para os governantes, sob pena das sanções previstas nesta Lei (Campos, 2009).

Oliveira (2009) afirma que o orçamento é um importante instrumento de política econômica, um elemento indispensável para o processo de planejamento. Para Silva e Weider (2012, p. 211), o planejamento questiona qual é o futuro das decisões que as organizações tomam. Dessa forma, o orçamento passou a ser peça estratégica e de controle financeiro, deixando compatível o ingresso de receitas ao volume de despesas em um determinado período de tempo. No entanto Almeida e Ferreira (2018, p. 27) ressaltam que “o orçamento brasileiro sempre enfatizou a despesa pública, deixando à previsão de receita o simples papel de formalizador dos recursos necessários ao financiamento das despesas orçamentárias”.

Outra questão importante é que diversas tentativas têm sido empreendidas para aperfeiçoar o planejamento (Rezende & Cunha, 2013). Entre elas destacam-se o emprego de maior racionalidade e funcionalidade aos instrumentos de planejamento, particularmente, ao Plano Plurianual (PPA), à Lei de Diretrizes Orçamentárias (LDO) e a Lei de Orçamento Anual (LOA) (Lima, 2005). E que, para o governo exercer suas múltiplas funções (alocativa, estabilizadora, distributiva), recursos são indispensáveis, e a principal forma de obtê-los é por meio da tributação.

Na Constituição Federal de 1988 as regras referentes ao orçamento, abrangem, a tributação. Por meio da tributação, busca-se arrecadar o tributo que também é entendido como sendo a “receita derivada que o Poder Público, com base no seu poder fiscal, arrecada, segundo o previsto em lei, do patrimônio particular do contribuinte com o escopo de atender às despesas da administração e custear os serviços públicos” (Diniz, 1998, p. 635).

O tributo tem por objeto uma obrigação jurídica de natureza pecuniária, ou seja, consiste na transferência de uma quantia em dinheiro para o Estado ou ente por ele autorizado, tal como no caso do Imposto de Renda (IR), Imposto sobre a Propriedade Territorial Urbana (IPTU) ou ainda, o Imposto sobre a Propriedade de Veículos Automotores (IPVA), sendo distinto da sanção por ato ilícito (Amaro, 2003).

Neste cenário, podemos inferir que, a previsão de receitas envolve o uso de técnicas analíticas para projetar a quantidade de recursos financeiros disponíveis num determinado tempo futuro. No setor público, as receitas vêm dos impostos, das taxas, das contribuições sociais e

econômicas ou de transferências intergovernamentais, além daquelas relativas à atividade governamental empresarial, ou seja, recursos diretamente arrecadados dos órgãos da administração direta e indireta do ente da federação (Almeida & Ferreira, 2018). A atividade de previsão, por sua vez, tenta identificar o relacionamento entre os fatores que afetam as receitas (alíquotas tributárias, variáveis macroeconômicas, volume de vendas a varejo etc.) e a arrecadação governamental (tributos e outras receitas). A habilidade de projetar, de forma precisa, os recursos futuros é bastante crítica, na medida em que se objetiva, entre outras coisas, evitar desequilíbrios orçamentários, objeto de preocupação constante das autoridades governamentais.

Como visto anteriormente, a Lei Complementar n. 101, de 04/05/2000 (LRF) obriga aos administradores públicos saberem com antecedência os resultados da arrecadação dos tributos que administram. Para isso, contam com diversos mecanismos e procedimentos assegurados pela referida lei em seus artigos 11 e 12.

Art. 11 - Constituem requisitos essenciais da responsabilidade na gestão fiscal a instituição, previsão e efetiva arrecadação de todos os tributos da competência constitucional do ente da federação; e

Art. 12 - As previsões de receitas observarão as normas técnicas e legais, considerarão os efeitos das alterações na legislação, da variação do índice de preços, do crescimento econômico ou de qualquer outro fator relevante e serão acompanhadas de demonstrativo de sua evolução nos últimos três anos (BRASIL. Lei complementar n. 101, de 04 de maio de 2000/LRF).

Neste sentido, fica evidente a relevância que os setores de previsão passaram a ter nas decisões estratégicas dos governos. No estado do Paraná, há poucos estudos relatando a aplicação de técnicas de previsão na arrecadação tributária. Os estudos de Liebel (2004, 2005) e Clemente e Clemente (2011) sobre a modelagem da arrecadação de ICMS do estado do Paraná destacam a importância da arrecadação de ICMS para o orçamento do governo do Paraná e a necessidade de previsões para o planejamento das despesas.

Liebel (2004) desenvolveu um trabalho voltado para a previsão do montante de ICMS no Paraná no ano de 2005. Por meio de metodologia Box e Jenkins, realizou previsões a partir de modelos univariados, dentre eles os modelos da classe ARIMA, especificamente para o ICMS do estado do Paraná. Seu objetivo foi testar vários métodos de previsão desde o mais simples (Regressão Linear Simples) ao mais complexo (SARIMA), a fim de determinar qual deles mais se adequa à previsão de ICMS. Dos resultados obtidos, foi possível observar que a complexidade do modelo não resulta necessariamente na melhor previsão, haja visto que o modelo escolhido para a previsão dos dados foi o de Holt Winters Aditivo e o método de suavização Exponencial de Holt com erro médio percentual de 0,4% e -0,7%, respectivamente.

Clemente e Clemente (2011), buscaram prever a arrecadação de ICMS para o período de agosto de 2011 a julho de 2012. Para tanto, aplicaram a metodologia Box e Jenkins para modelar a série mensal de ICMS de janeiro de 2000 a julho de 2011, a qual foi diferenciada para tornar-se estacionária. A partir da análise de resíduo e dos testes de significância dos parâmetros, elegeu-se a identificação SARIMA (2,1,0)(0,1,1)₁₂ para previsão. De acordo com este modelo, a arrecadação de ICMS crescerá 13,2% entre julho de 2011 e janeiro de 2012, diminuirá 5% entre janeiro e fevereiro de 2012, e aumentará 6,7% entre fevereiro e julho de 2012.

3. BASE DE DADOS E METODOLOGIA

Os dados coletados são referentes à arrecadação do estado do Paraná, ao índice de atividade econômica do estado e do índice de inflação. A *proxy* utilizada para a arrecadação do estado

são referentes aos impostos de competência do estado: ICMS (imposto sobre a circulação de mercadorias), IPVA (imposto sobre a propriedade de veículos automotores), ITCMD (imposto sobre transmissão causa mortis e doação) e outras receitas administrativas em mil reais. Os dados de atividade econômica são referentes a *proxy* do Índice de atividade econômica do Banco Central IBC-Br para o estado do Paraná, e a *proxy* para inflação é o IPCA (Índice Nacional de Preços ao Consumidor Amplo) em percentual ao mês.

A fonte dos dados referentes a arrecadação foram obtidos junto ao IPEADATA, a série do IBC-Br para o estado do Paraná tem como fonte o Banco Central, e os dados para o IPCA foram obtidos no IBGE (Instituto Brasileiro de Geografia e Estatística). O período das séries perfazem os meses de dezembro de 2003 a abril de 2021.

As séries de arrecadação foram convertidas em uma única série, composta da soma destas, e convertida em dados reais. Além dos dados de arrecadação serem deflacionados pelo IPCA, o mesmo ainda foi acumulado para 12 meses, possibilitando desta forma uma suavização dos valores. Os dados do IPCA e do IBC-Br foram acumulados também para 12 meses, apresentando respectivamente uma taxa de inflação anual e um índice para 12 meses.

Em relação ao modelo estatístico utilizado, é importante primeiramente analisar a existência ou não de uma correlação cruzada entre as variáveis analisadas. Para tanto é necessário a utilização da função de correlação cruzada (*cross correlation function* - CCF) entre as séries, que demonstrará os *lags* k 's em que as variáveis são correlacionados.

É importante ainda salientar que a CCF deve ser aplicada a séries temporais estacionárias, sendo necessário a indução de normalidade caso as séries não o sejam, através de diferenciação. Mas na ocorrência de fortes correlações de curto prazo, faz-se necessário ainda a filtragem dos dados através da metodologia *pre-whitening*, com o objetivo de se evitar a ocorrência de correlações espúrias entre os dados, aplicando-se após a filtragem a CCF. No caso da ocorrência de correlação espúria, a CCF com filtro *pre-whitening* não indicará a existência de correlação, ao mesmo tempo que a CCF sem filtro demonstrará (Chatfield & Xing, 2019).

Uma vez identificado a ocorrência de correlação cruzada entre as séries estacionárias, eliminando-se a possibilidade de existência de correlações espúrias, é possível transcrever o modelo de regressão como:

$$(1) \quad Y_t = \beta_0 + \beta_1 X_{t-k} + \epsilon_t$$

sendo: Y é a variável explicada, X são as variáveis explicativas, que são independentes, identicamente distribuídas e com variância σ_X^2 , e ϵ_t é o ruído branco com variância σ_ϵ^2 e independente de X (Cryer & Chan, 2008).

Mas vale ressaltar que o modelo da equação (1) é relativamente restrito a condições de que as variáveis X e Y sejam aleatórias (Morettin & Toloi, 2004; Cryer & Chan, 2008). Sendo que para séries temporais estacionárias, as variáveis X e Y são geralmente autocorrelacionadas, tornando o termo de erro autocorrelacionado e não independente em relação a X . Desta maneira, segundo Cryer e Chan (2008), o ideal seria considerar o modelo de regressão acima acrescido de um modelo de séries temporais do tipo ARIMA (*Autoregressive Integrated Moving Average*), que viabilize a modelagem das autocorrelações existentes como:

$$(2) \quad Y_t = \beta_0 + \beta_1 X_{t-k} + Z_t$$

sendo: Z_t é um modelo de séries temporais do tipo ARIMA.

Segundo Shumway e Stoffer (2011), o modelo da equação (2), que mescla um modelo de regressão a um modelo de séries temporais ARIMA, é denominado de modelo ARIMAX

(*Autoregressive Integrated Moving Average with Exogenous Inputs*). Desta forma, considerando que o objetivo do trabalho é desenvolver um modelo de previsão para a arrecadação do estado do Paraná, tem-se na equação (2) o modelo estatístico a ser utilizado. Considerando que Z_t na equação é um modelo de séries temporais do tipo ARIMA, que possibilitará desenvolver o modelo de previsão para o produto, faz-se necessário a compreensão destes modelos.

A metodologia de previsão ARIMA é composta de componentes autorregressivos $\phi(B)$, e de médias móveis $\theta(B)$ para uma dada variável Z_t , considerando ainda a ordem de diferenciação desta variável $(1 - B)^d Z_t$, dada a necessidade de que a mesma seja estacionária (Chatfield & Xing, 2019). Desta maneira, obtém-se o modelo multiplicativo conforme:

$$(3) \quad \phi(B)(1 - B)^d Z_t = \theta(B)\epsilon_t$$

sendo: $\phi(B) = (1 - \alpha_1 B - \dots - \alpha_p B^p)$, $\theta(B) = (1 + \beta_1 B + \dots + \beta_q B^q)$, sendo B o operador de retardo.

Na equação (3) ainda é possível observar a ordem p para autorregressivo, d para diferenciação e q para médias móveis, representando assim o modelo ARIMA (p, d, q) .

Mas o modelo de previsão utilizado no presente estudo é baseado no modelo proposto por Box e Jenkins (1976), para o caso de sazonalidade da série, quando os dados apresentam um comportamento periódico no transcorrer do tempo, sendo este o modelo SARIMA, uma derivação do modelo ARIMA com sazonalidade. Desta maneira, o modelo sazonal multiplicativo pode ser representado como:

$$(4) \quad \phi(B)\Phi(B^s)W_t = \theta(B)\Theta(B^s)\epsilon_t$$

onde: $\phi(B) = (1 - \alpha_1 B - \dots - \alpha_p B^p)$, $\Phi(B^s) = (1 - \phi_1 B^s - \dots - \phi_P B^{Ps})$, $W_t = \nabla^d \nabla_s^D Z_t$, $\theta(B) = (1 + \beta_1 B + \dots + \beta_q B^q)$, $\Theta(B^s) = (1 + \theta_1 B^s + \dots + \theta_Q B^{Qs})$, sendo B o operador de retardo.

A equação (4) especifica o modelo SARIMA genérico de ordem $(p, d, q) \times (P, D, Q)_s$, sendo p , d e q as ordens autorregressivas, de diferenciação e de médias móveis respectivamente do ARIMA. Enquanto P , D e Q representam as mesmas ordens sazonais, representadas pelo subscrito s , que indica a frequência da sazonalidade, sendo 12 para meses, 4 para trimestres e assim sucessivamente (Chatfield & Xing, 2019).

Para se determinar qual o modelo SARIMA mais adequado as séries, faz-se necessário inicialmente formular o modelo, estimar e verificar, sendo o passo de formulação o mais dispendioso, dada sua complexidade. Desta forma, para se formular o modelo é fundamental observar as funções de autocorrelação (*autocorrelation function* - ACF) e autocorrelação parcial (*partial autocorrelation function* - *Partial ACF*) em relação a variável analisada (Chatfield & Xing, 2019).

Decaimentos exponenciais e oscilatórios na ACF e quedas abruptas a partir de determinado *lag k* na *parcial ACF*, indicam um modelo autorregressivo de ordem $p = k$, enquanto decaimentos exponenciais e oscilatórios na *parcial ACF* e quedas abruptas a partir de determinado *lag k* na ACF, indicam um modelo de médias móveis de ordem $q = k$. Já para o caso de um modelo ARMA, é necessário que ocorram quedas abruptas a partir de determinados *lags* em ambas as funções, sendo que o *lag k* na ACF determinará a ordem de médias móveis q , e o *lag k* na *parcial ACF* a ordem do autorregressivo p (Cryer & Chan, 2008).

É possível ainda se utilizar da matriz de função de autocorrelação estendida (*extended autocorrelation function* - EACF), que facilita a indicação do modelo de séries temporais, principalmente no caso de séries que necessitem tanto de parâmetros autorregressivos quanto de médias móveis.

Já para o caso de um modelo SARIMAX, a ACF, a *parcial* ACF e a função de autocorrelação estendida EACF, devem ser calculadas baseadas nos resíduos de um modelo de regressão linear, possibilitando assim identificar o modelo de previsão SARIMA mais adequado a regressão (Cryer & Chan, 2008).

É importante ainda salientar a necessidade da utilização das análises de intervenção no modelo proposto, dada a existência de *outliers* nas séries utilizadas. Viabilizando assim, solucionar o problema da distribuição de caudas pesadas. Conforme observado por Chang, Tiao e Chen (1988), as distorções decorrentes de *outliers* podem ser modeladas através da utilização de modelos de intervenção, sendo que existem dois tipos distintos de *outliers*, os aditivos e os inovativos.

Os *outliers* aditivos são aqueles que afetam apenas uma observação, enquanto os inovativos representam perturbações que se manifestam no transcorrer do tempo a partir de determinado período t .

É importante ainda salientar que o tratamento dos *outliers* deve ocorrer através da utilização de variáveis *dummies*, na qual os *outliers* apresentarão valor 1, enquanto as outras observações 0. O modelo de intervenção segue como:

$$(5) \quad Y_t = \sum_{i=1}^K v_i(B)x_{i,t} + Z_t$$

sendo: Y_t é a variável dependente, K o número de intervenções na série, $v_i(B)$ é o valor da função de transferência, x a variável binária e Z_t um modelo SARIMA.

Desta forma é possível determinar o modelo finalizado SARIMAX a ser aplicado no presente estudo conforme:

$$(6) \quad Y_t = \beta_0 + \beta_1 X_{t-k} + \beta_2 D_t + Z_t$$

sendo: D_t são as variáveis *dummies* das análises de intervenção.

Uma vez formulado e estimado o modelo SARIMAX faz-se necessário verificar a adequação do modelo, aplicando-se os testes de Box-Pierce e Ljung-Box para avaliar a independência dos resíduos, assim como os testes de Jarque-Bera e Shapiro-Wilk para avaliar a normalidade dos resíduos.

Já em relação ao caso de um comparativo entre modelos distintos, é necessário uma avaliação dos resultados obtidos com o critério de informação Akaike - AIC e o valor da variância σ^2 entre os modelos, sendo que o modelo que apresentar o valor AIC mais negativo com variância menor, apresentar-se-á mais adequado.

4. ANÁLISE E DISCUSSÃO DOS DADOS

Conforme discutido na seção anterior, para que seja possível desenvolver o modelo SARIMAX é necessário inicialmente observar a existência de correlação cruzada entre as variáveis discutidas. Desta forma, a relação existente entre a arrecadação do estado do Paraná e o IBC-BR e o IPCA, podem ser comprovados quando aplicada a processos estacionários a função de correlação cruzada (*cross correlation function* - CCF). Sendo necessário ainda, conforme discutido na seção de metodologia, a aplicação do filtro *pre-whitening* as séries, que uma vez aplicada a CCF, demonstrará a existência ou não de correlação não espúria as variáveis analisadas.

Como as três séries são processos estacionários homogêneos, obteve-se desta forma os resultados de correlação cruzada na Figura 01, que demonstra nos primeiros gráfico a esquerda a CCF das variáveis sem filtragem, enquanto os segundos gráficos a direita tem-se a CCF com a aplicação do filtro *pre-whitening* aos dados. A indicação de correlação cruzada em todas as CCF indica a existência de correlação não espúria entre as variáveis, viabilizando desta forma o desenvolvimento do modelo.

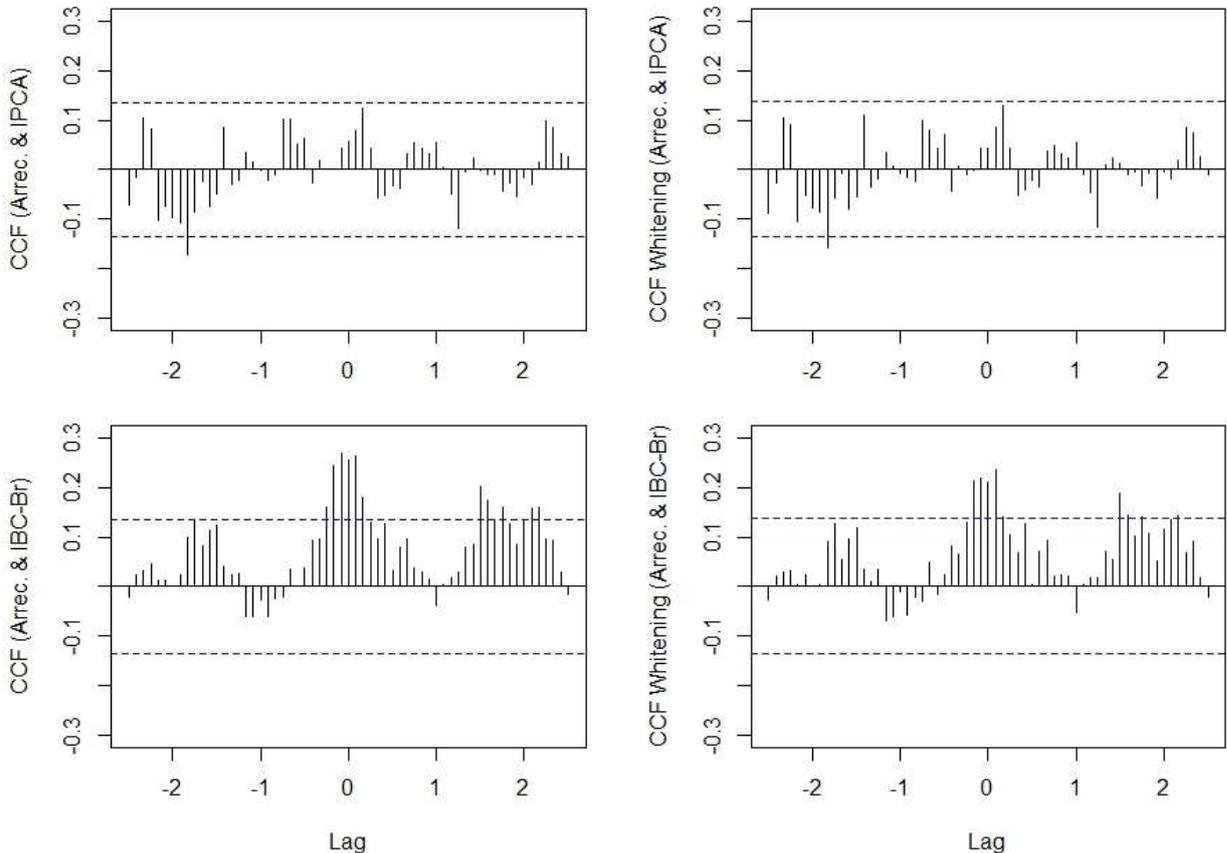


Figura 01 – Função de Correlação Cruzada (Arrecadação vs IPCA e IBC-Br)
Fonte: Elaborado a partir de Banco Central e IBGE 2021.

Uma vez analisada a existência de correlação cruzada entre as variáveis, faz-se necessário formular o modelo de séries temporais. Considerando que o modelo SARIMAX deve ser formulado mediante os resíduos de um modelo de regressão linear, é necessário estimar um modelo de mínimos quadrados ordinários (*ordinary least squares* - OLS) conforme o modelo teórico apresentado na seção 3.

A Tabela 01 demonstra os resultados do modelo de regressão linear, que apesar da indicação correta dos parâmetros e da significância, apresenta autocorrelação não modelada nos resíduos, conforme indicado pelos testes de hipótese de Box-Pierce e Box-Ljung, não confirmando a hipótese nula de independência dos resíduos.

Variáveis	Coefficientes	Erro Padrão	Valor - t
Intercepto	0,001763	0,000901	1,9561 *
IPCA	0,002957	0,008791	0,3364
IBC-Br	0,766091	0,207367	3,6944 ***

Variável dependente: Arrecadação do Estado do Paraná acumulado de 12 meses.
Significância: 1% ***, 5% **, 10% *

Tabela 01 – Estimativa Coeficientes OLS

Através das funções de autocorrelação (*autocorrelation function* - ACF) e autocorrelação parcial (*partial autocorrelation function* - Partial ACF), em relação aos resíduos do modelo linear apresentado na Tabela 01, é possível observar a indicação do modelo SARIMAX a ser desenvolvido. A Figura 02 demonstra apenas a existência de autocorrelação no lag 12 tanto na ACF quanto na parcial ACF, indicando desta forma que o SARIMA não apresenta ordens autorregressivas e médias móveis de curto prazo, não especificando de forma clara as ordens de longo prazo.

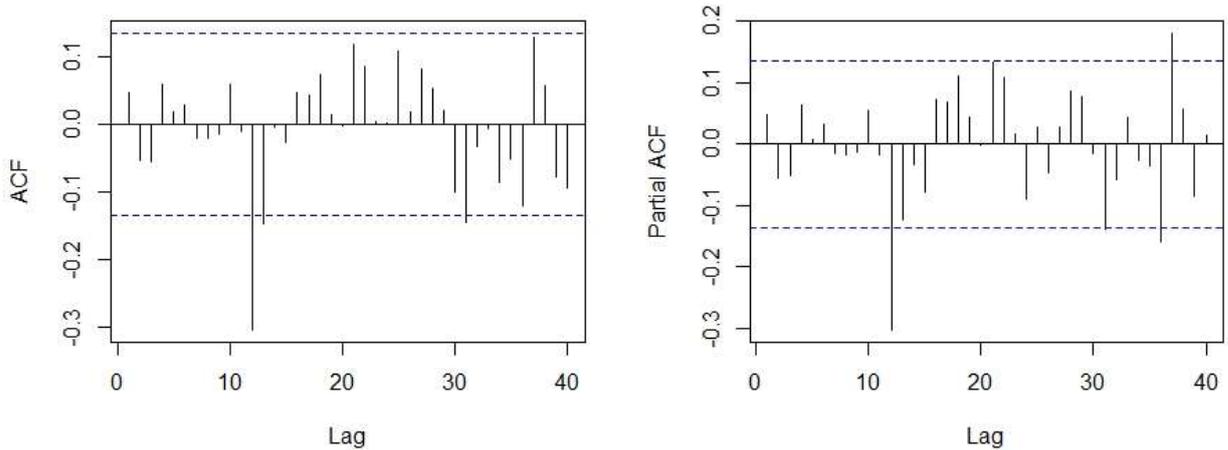


Figura 02 – Função de Autocorrelação e Autocorrelação Parcial (Resíduos OLS)
 Fonte: Elaborado a partir de Banco Central e IBGE 2021.

Mas com a utilização da matriz EACF, conforme especificado na Figura 3, é possível que se obtenha uma indicação mais clara. A matriz especifica nas linhas a ordem do modelo autorregressivo e nas colunas o modelo de médias móveis. O método demonstra em uma matriz preenchida com X e 0's, o vértice estilizado de um triângulo da direita para a esquerda e de baixo para cima, formado por 0's (Cryer & Chan, 2008). Neste caso foi possível obter a indicação de um modelo SARIMA (0,1,0)(0,0,1)₁₂.

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	o	o	o	o	o	o	o	o	o	o	o	x	o	o
1	x	o	o	o	o	o	o	o	o	o	o	x	o	o
2	x	x	o	o	o	o	o	o	o	o	o	x	x	o
3	x	x	x	o	o	o	o	o	o	o	o	x	x	o
4	o	x	x	x	o	o	o	o	o	o	o	x	o	o
5	o	x	x	x	o	o	o	o	o	o	o	x	o	o
6	x	x	x	o	o	o	o	o	o	o	o	x	x	o
7	x	x	x	o	o	o	o	o	o	o	o	x	x	o

Figura 03 – Matriz EACF da Amostra dos Resíduos do Modelo Linear
 Fonte: Elaborado a partir de Banco Central e IBGE 2021.

Um vez determinado o modelo SARIMA (0,1,0)(0,0,1)₁₂, é possível estimar o modelo SARIMAX, aos moldes do modelo de regressão abaixo:

$$(7) \quad \ln Arred_t + \nabla Z_t = \delta_0 + \delta_1 \ln \pi_t + \delta_2 \ln IBC_t + (1 + \theta_1 B^{12})\epsilon_t$$

sendo: $\nabla Z_t = (1 - B)Z_t = Z_t - Z_{t-1}$, $Arred_t$ a arrecadação do estado do Paraná, π_t o índice de inflação IPCA e IBC_t o índice de atividade econômica IBC-Br.

A representação do modelo na forma multiplicativa é eventualmente preferível, dado a possibilidade de representar uma grande quantidade de parâmetros de forma resumida. Mas é possível ainda representar a equação (7) na forma de equação das diferenças, que é relativamente mais extensa dado o número de parâmetros.

$$(8) \quad \ln Arred_t = \delta_0 + \delta_1 \ln \pi_t + \delta_2 \ln IBC_t - (Z_t - Z_{t-1}) + \epsilon_t + \theta_1 \epsilon_{t-12}$$

A equação (8) demonstra mais claramente que a variável explicada $Arred_t$ não está em função unicamente das variáveis exógenas π_t e IBC_t , mas é determinada também por um termo de médias móveis sazonal ϵ_{t-12} , representando desta forma o modelo SARIMA.

É possível assim afirmar que o acréscimo do modelo SARIMA ao modelo de regressão, tem como objetivo principal, modelar as autocorrelações residuais existentes no modelo linear, caracterizando assim o modelo SARIMAX.

A estimação da equação (8), representada pela Tabela 02, demonstra que o coeficiente sazonal de médias móveis e da variável exógena IBC-Br, apresentaram significância ao nível de 1%, enquanto o parâmetros da variável IPCA não apresentou significância. Mas quando se avalia a relação demonstrada, ambas as variáveis exógenas demonstraram uma relação diretamente proporcional conforme o esperado.

Variáveis	Coefficientes	Erro Padrão	Valor - t
Constante	0,0017	0,0006	2,8333 ***
SMA	-0,4172	0,0789	-5,2877 ***
IPCA	0,0059	0,0092	0,6413
IBC-Br	0,8053	0,1811	4,4467 ***
AIC	-1.301,25		
σ^2	1,080E-04		

Variável dependente: Arrecadação do Estado do Paraná acumulado de 12 meses.

Significância: 1% ***, 5% **, 10% *

Tabela 02 – Estimativa Coeficientes SARIMAX (0,1,0)(0,0,1)₁₂

Aplicando-se os testes de Box-Pierce e Box-Ljung para avaliar a independência dos resíduos, obteve-se em ambos os testes a confirmação da hipótese nula de independência, solucionando assim o problema de autocorrelação dos resíduos. Já em relação aos testes de Jarque-Bera e Shapiro-Wilk para determinação da normalidade dos resíduos, obteve-se a não confirmação da hipótese nula de normalidade, dado a existência de caudas pesadas nos resíduos.

Para solucionar este problema, é necessário que seja aplicado ao modelo acima, análises de intervenção em relação aos *outliers* existentes, conforme discutido na sessão de metodologia. Desta forma foram identificados 18 *outliers*, sendo 3 deles aditivos e 15 inovativo.

Dentre os aditivos tem-se: dezembro de 2013, junho de 2014 e dezembro de 2015. E em relação aos *outliers* inovativos tem-se: junho e julho de 2013, fevereiro a maio de 2015, janeiro e fevereiro de 2016, abril e maio de 2016, janeiro e fevereiro de 2017, janeiro e fevereiro de 2018 e julho de 2020.

A Tabela 03 demonstra a estimativa dos coeficientes do modelo SARIMA (0,1,0)(0,0,1)₁₂, com a inclusão das análises de intervenção. Como pode ser observado todos os coeficientes apresentam significância a 1%, a exceção do IPCA que apresentou significância a 10%. Em relação aos parâmetros das variáveis exógenas, obteve-se também um valor positivo em ambas as variáveis, em linha com a expectativa.

Variáveis	Coefficientes	Erro Padrão	Valor - t
Constante	0,0032	0,0006	5,3333 ***
SMA	-0,9999	0,0860	-11,6267 ***
IPCA	0,0096	0,0052	1,8462 *
IBC-Br	0,5077	0,1479	3,4327 ***
Outlier A1 (12/2013)	0,0416	0,0037	11,2432 ***
Outlier A2 (06/2014)	-0,0433	0,0050	-8,6600 ***
Outlier A3 (12/2015)	-0,0396	0,0039	-10,1538 ***
Outlier I1 (06/2013)	0,0430	0,0048	8,9583 ***
Outlier I2 (07/2013)	-0,0441	0,0048	-9,1875 ***
Outlier I3 (02/2015)	-0,0240	0,0040	-6,0000 ***
Outlier I4 (03/2015)	0,0112	0,0031	3,6129 ***
Outlier I5 (04/2015)	0,0455	0,0060	7,5833 ***
Outlier I6 (05/2015)	-0,0270	0,0051	-5,2941 ***
Outlier I7 (01/2016)	0,0293	0,0051	5,7451 ***
Outlier I8 (02/2016)	-0,0204	0,0062	-3,2903 ***
Outlier I9 (04/2016)	-0,0497	0,0060	-8,2833 ***
Outlier I10 (05/2016)	0,0250	0,0053	4,7170 ***
Outlier I11 (01/2017)	0,0596	0,0068	8,7647 ***
Outlier I12 (02/2017)	-0,0471	0,0067	-7,0299 ***
Outlier I13 (01/2018)	-0,0389	0,0053	-7,3396 ***
Outlier I14 (02/2018)	0,0314	0,0053	5,9245 ***
Outlier I15 (07/2020)	0,0085	0,0032	2,6563 ***
AIC	-1.573,42		
σ^2	2,301E-05		

Variável dependente: Arrecadação do Estado do Paraná acumulado de 12 meses.

Significância: 1% ***, 5% **, 10% *

Tabela 03 – Estimativa Coeficientes SARIMAX (0,1,0)(0,0,1)₁₂ com Intervenção

Em relação a verificação do modelo, os testes de independência dos resíduos de Box-Pierce e Box-Ljung, confirmaram a hipótese nula de independência, apresentando os resultados de 0,8395 e 0,7314 respectivamente de seus valores-p, indicando, desta forma, a não existência de autocorrelação residual.

Em relação aos testes de normalidade dos resíduos, os resultados para os testes de Jarque-Bera e Shapiro-Wilk apresentaram valores-p de 0,1294 e 0,1592, respectivamente, confirmando a hipótese nula de normalidade. Na Figura 04 é possível observar o histograma dos resíduos em relação a curva normal e o gráfico normal *q-q plot*, que indicam, para este caso, a solução do problema de cauda pesada nos resíduos.

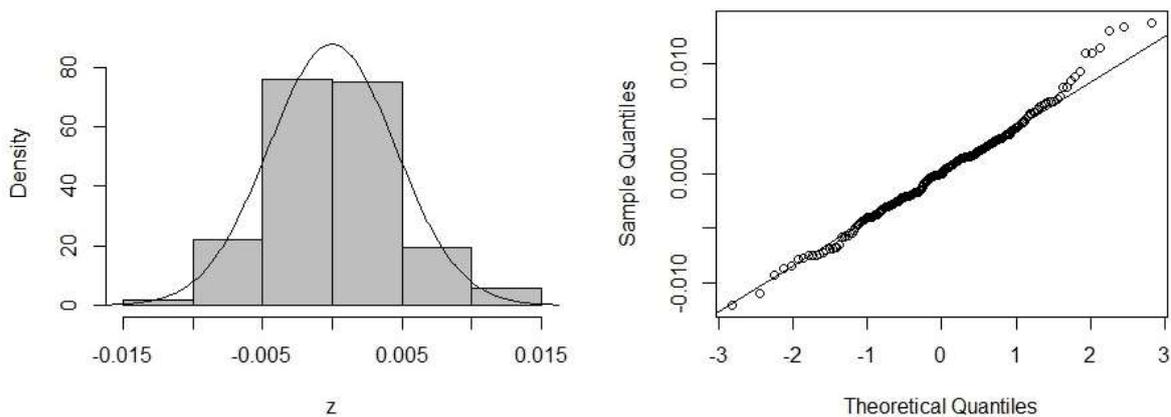


Figura 04 – Histograma e Normal *Q-Q Plot* SARIMA (0,1,0)(0,0,1)₁₂ com Intervenção
 Fonte: Elaborado a partir de Banco Central e IBGE 2021.

Uma vez avaliado o modelo SARIMAX proposto, é relevante comparar o mesmo com um modelo de séries temporais SARIMA com as mesmas especificações, com o objetivo de observar se o modelo SARIMAX incluso das variáveis IBC-Br e IPCA para previsão da arrecadação do estado do Paraná acumulada de 12 meses é realmente mais acurado.

A Tabela 04 demonstra a estimação dos coeficientes para o SARIMA (0,1,0)(0,0,1)₁₂, seguindo as especificações indicadas do modelo anterior proposto. Conforme pode ser observado, os coeficientes demonstraram todos significância a 1%, indicando um bom ajuste do modelo. Mas quando se compara os resultados dos critérios de informação Akaike - AIC e a variância σ^2 entre os modelos, percebe-se que o AIC é maior em módulo em quase 200 unidades do modelo SARIMAX em relação ao SARIMA, enquanto a variância do SARIMAX é menor que a do SARIMA, indicando-se desta forma uma melhor adequação do modelo proposto frente ao modelo que desconsidera as variáveis exógenas.

VARIÁVEIS	COEFICIENTES	ERRO PADRÃO	VALOR - T
Constante	0,0035	0,0005	7,0000 ***
SMA	-0,3337	0,0689	-4,8433 ***
AIC	-1.285,87		
σ^2	1,179E-04		

Variável dependente: Arrecadação do Estado do Paraná acumulado de 12 meses.
 Significância: 1% ***, 5% **, 10% *

Tabela 04 – Estimativa Coeficientes SARIMA (0,1,0)(0,0,1)₁₂

É possível ainda observar com a Figura 05, a diferença de comportamento dos resíduos dos dois modelos, o SARIMAX e o SARIMA. À esquerda da figura tem-se o comportamento dos resíduos para o modelo SARIMAX e a direita os resíduos do modelo SARIMA, possibilitando perceber um comportamento muito mais homogêneo do modelo proposto em relação ao modelo de univariado.

A média dos erros para o processo aleatório do modelo SARIMAX é claramente observada como oscilando em torno de zero, situação não tão clara de ser percebida para o modelo SARIMA, além dos *outliers* e da variância não constante que este modelo apresenta. Indicando, desta forma, novamente a preferência em se utilizar o modelo SARIMAX.

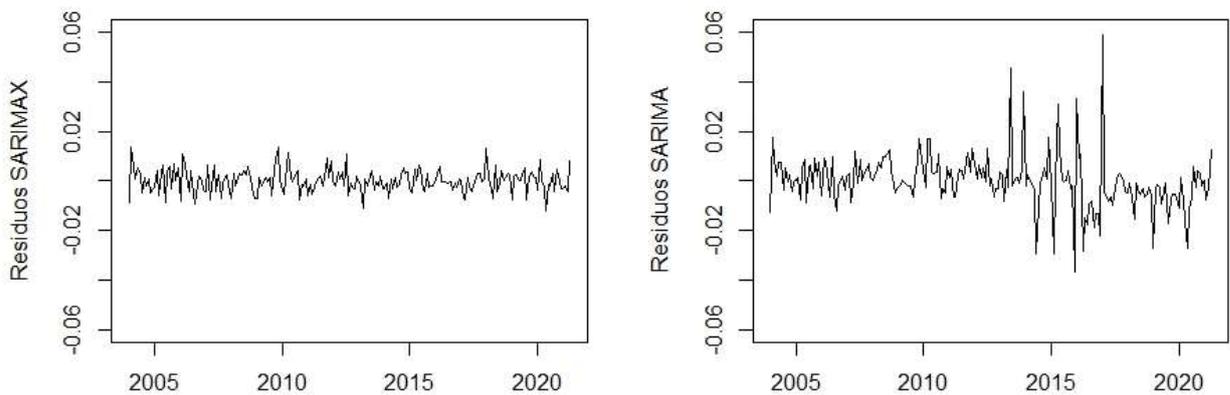


Figura 05 – Resíduos dos Modelos SARIMAX e SARIMA
 Fonte: Elaborado a partir de Banco Central e IBGE 2021.

Uma vez avaliado a adequação do SARIMAX, é possível partir agora para confirmação visual do modelo e do processo de previsão. A Figuras 06 demonstra o comportamento da série original em relação à série estimada pelo modelo SARIMAX, para a arrecadação do Estado do Paraná acumulada de 12 meses.

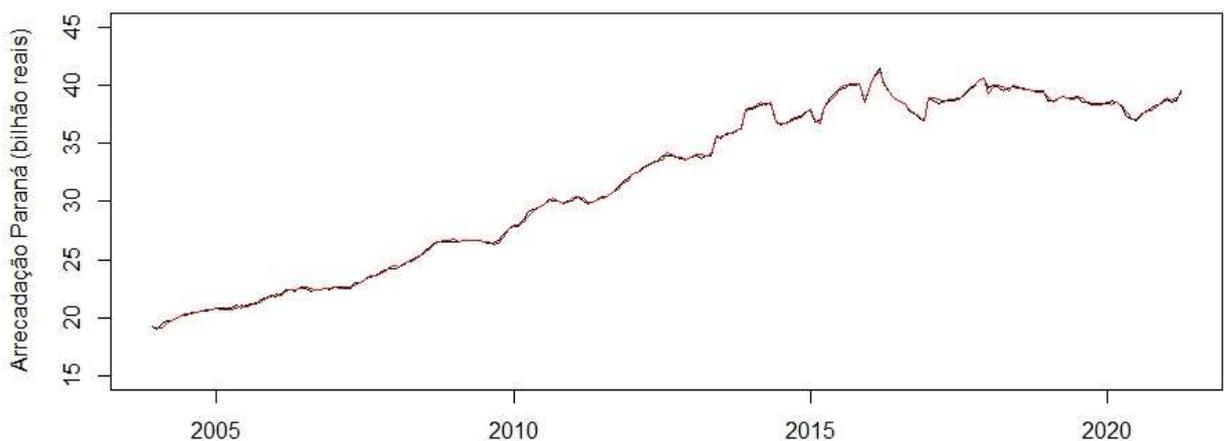


Figura 06 – Arrecadação do Estado do Paraná (Efetivo e Estimado)
 Fonte: Elaborado a partir de Banco Central e IBGE 2021

A Figura 06 indica o comportamento da arrecadação do estado do Paraná acumulada de 12 meses, dado pela linha escura, e a arrecadação, linha vermelha, demonstrando uma boa adequação do modelo à série original, considerando que as séries praticamente se sobrepõem.

Já a previsão para a arrecadação do estado do Paraná acumulado de 12 meses em bilhões de reais é apresentada na Figura 07, com a indicação da projeção da arrecadação em linha azulada para os próximos 12 meses, iniciando-se no mês de maio de 2021 a abril de 2022, dado que o último mês da série é abril de 2020.

A projeção da arrecadação que se inicia em maio de 2021 indica um valor de 39,352 bilhões de reais, alcançando no mês de abril de 2022 o montante de 40,568 bilhões de reais. As áreas sombreadas na figura indicam o desvio padrão para a previsão, sendo o sombreado mais claro condizente com um desvio de 80% da normal em relação a projeção, enquanto o sombreado mais escuro especifica 95% da normal, ou seja, construído a partir de aproximadamente 2 desvios padrão.

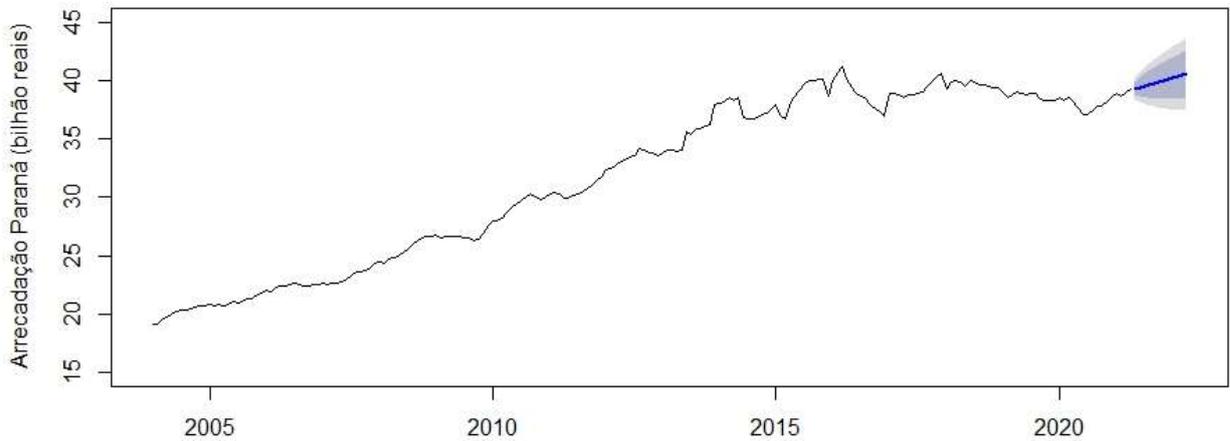


Figura 07 – Projeção da Arrecadação do Estado do Paraná Acumulado de 12 meses
Fonte: Elaborado a partir de Banco Central e IBGE 2021.

Para o primeiro desvio de 80%, em relação ao primeiro período de previsão, obteve-se 38,769 e 39,935 bilhões de reais para os limites inferior e superior respectivamente, enquanto para o último período de previsão obteve-se 38,515 e 42,621 bilhões de reais para os limites. Já em relação ao segundo desvio, obteve-se 38,460 e 40,244 bilhões de reais para os limites inferior e superior respectivamente, e para o último período da séries de previsão obteve-se 37,428 e 43,708 bilhões de reais para os limites.

Desta forma, é possível afirmar, conforme o modelo de previsão desenvolvido no presente trabalho, que a arrecadação do estado do Paraná pode apresentar no prazo de 1 ano a frente, um crescimento de 2,57% na média da projeção, com uma oscilação máxima de -5,37% a 10,51% para o período de 12 meses de maio de 2021 a abril de 2022.

Mas é possível ainda retroceder 12 meses ao período estimado para que seja possível comparar os resultados da previsão com os últimos 12 meses de dados efetivos, conforme indicado na Figura 08.

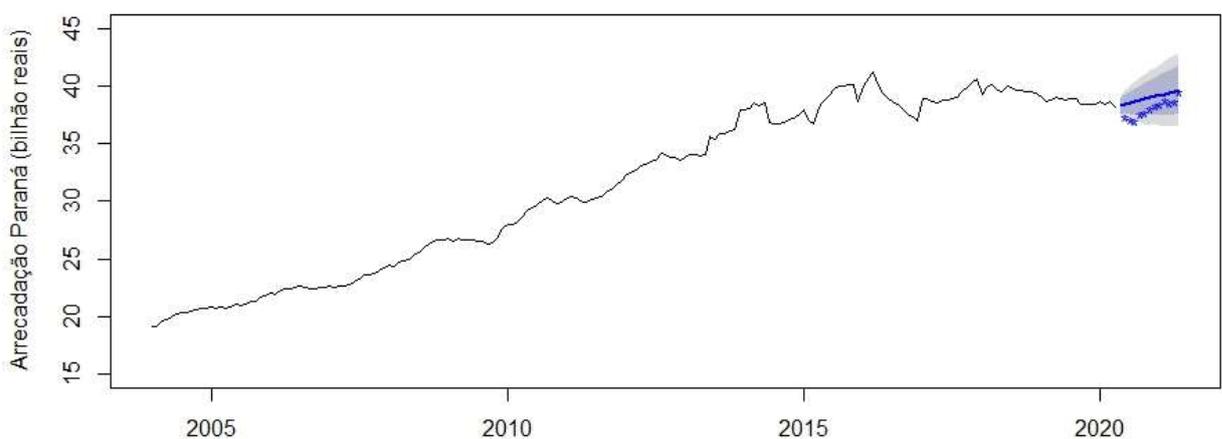


Figura 08 – Projeção da Arrecadação do Estado do Paraná Acumulado de 12 meses
Fonte: Elaborado a partir de Banco Central e IBGE 2021.

A Figura 08 demonstra, além do modelo estatístico para a arrecadação do estado do Paraná e das previsões, os 12 meses subsequentes efetivados de arrecadação, indicados como asteriscos

azulados se sobrepondo aos desvios da previsão. Dentre os 12 meses de arrecadação acumuladas de 12 meses efetivos para o estado, observou-se que apenas o terceiro mês da série, o mês de julho de 2020, o valor ficou fora da projeção, com o limite inferior indicando 37,022 bilhões de reais e o valor efetivo indicando 36,969 bilhões.

Apesar deste desvio, os valores são relativamente próximos, indicando uma diferença de apenas 53 milhões de reais, condizente com 0,14% do total. Valendo ainda ressaltar que a queda da arrecadação ocorreu em um cenário de grande instabilidade em decorrência da crise sanitária do Covid-19. Já em relação aos outros valores na sequência deste, permaneceram dentro da margem de erro da projeção, indicando uma boa acuracidade de previsão do modelo proposto para a arrecadação do estado do Paraná.

CONCLUSÃO

Uma vez avaliado o modelo SARIMAX, é possível perceber a sua eficiência frente ao modelo SARIMA dado o comportamento dos resíduos e das diferenças indicadas a favor do modelo SARIMAX em relação ao critério de informação AIC e ao desvio padrão. Mas, além dos resultados técnicos favoráveis ao modelo SARIMAX, demonstrou-se também a acuracidade do modelo dado que os valores nominais para 12 períodos anteriores se ajustaram dentro da margem de risco prevista.

Percebe-se desta maneira, que uma gestão que busque fixar as despesas do Estado baseando-se no limite inferior da previsão ao invés da média da arrecadação, pode proporcionar uma redução expressiva do risco na gestão fiscal, possibilitando ainda a geração de caixa para possíveis cenários que venham a exceder os limites inferiores da previsão.

Conclui-se desta forma, que o método de séries temporais SARIMAX apresenta-se como um excelente instrumento para previsão da arrecadação do Paraná, auxiliando desta forma na gestão fiscal do estado, sobre o prisma da Lei de Responsabilidade Fiscal.

REFERÊNCIAS

- Aguilar, A. M. Gestão fiscal responsável. In: Castro, C. L. F. de, Gontijo C. R. B. & Amabile A. E. de N. (2012). *Dicionário de políticas públicas*. Barbacena: EdUEMG.
- Amaro, L. (2003). *Direito Tributário Brasileiro*. 9. ed. São Paulo: Saraiva, 2003.
- Barbosa, R. B.; Ferreira, R. T.; Silva, T. M. (2020). Previsão de variáveis macroeconômicas brasileiras usando modelos de séries temporais de alta dimensão. *Estud. Econ.* 50 (1), Jan-Mar 2020. Disponível em: <https://doi.org/10.1590/0101-41615013rrt>. Acesso em: 11 ago. 2021.
- BRASIL. Lei Complementar Federal n. 101, de 4 de maio de 2000. Lei de Responsabilidade Fiscal. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. *Diário Oficial da União*, Brasília, 5 de maio de 2000. Disponível em: http://www.planalto.gov.br/ccivil_03/Leis/LCP/Lcp101.htm. Acesso em: 10 ago. 2021.
- BRASIL. Constituição (1988). *Constituição da República Federativa do Brasil*, 1988. Brasília, Senado. Disponível em: <http://www.presidencia.gov.br>. Acesso em: 10 ago. 2021.
- Box, G. E. P. & Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*. 2 ed. San Francisco: Horden-Day.
- Campos, C. V. C. (2009). *Métodos econométricos aplicados à previsão da arrecadação de receitas federais*. Secretaria da Receita Federal do Brasil – 8º Prêmio Schöntag. Ribeirão Preto-SP. Disponível em: <https://repositorio.enap.gov.br/bitstream/1/4588/1/3%c2%ba%20lugar%20do%208%c2%ba%20Premio%20RFB.pdf>. Acesso em: 09 ago. 2021.
- Cardoso Jr., J. (Org.). (2011). *A reinvenção do planejamento governamental no Brasil*. Brasília: Ipea. v. 4, p. 517. (Diálogos para o Desenvolvimento). Disponível em: <http://repositorio.ipea.gov.br/bitstream/11058/3182/1/A%20reinven%C3%A7%C3%A3o%20do%20planejamento%20governamental%20no%20Brasil.pdf>. Acesso em: 09 ago. 2021.

- Chatfield, C.; Xing, H. (2019). *The Analysis of Time Series: An Introduction with R*. 7 ed. Boca Raton: CRC Press.
- Chang, I.; Tiao, G.C. & Chen, C. (1988). Estimation of Time Series Parameters in the Presence of Outliers. *Technometrics*, 30 (2), p. 193-204.
- Cryer, J. & Chan, K. (2008). *Time Series Analysis: With Applications in R*. 2 ed. New York: Springer.
- Dagnino, R. P. (2014). *Planejamento estratégico governamental*. 3. ed. rev. atual. Florianópolis: Departamento de Ciências da Administração-UFSC. Brasília: CAPES: UAB, 2014.
- Ferreira, R., Bierens, H. & Castelar, I. (2005). Forecasting quarterly Brazilian GDP growth rate with linear and nonlinear diffusion index models. *Economia, Selecta*, Brasília-DF, v.6, n.3, p.261-292, 2005. Disponível em: http://www.anpec.org.br/revista/vol6/vol6n3p261_292.pdf. Acesso em: 11 ago. 2021.
- Kjellin, M. (2020). *Modelagem e identificação de séries temporais aplicados ao mercado financeiro*. Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro Tecnológico, Graduação em Engenharia de Controle e Automação, Florianópolis. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/204795/TCC%20pdfa.pdf?sequence=3&isAllowed=y>. Acesso em: 11 ago. 2021.
- Liebel, M. J. (2004). *Previsão de receitas tributárias - o caso do ICMS no estado do Paraná*. Dissertação (Mestrado Profissionalizante em Engenharia) - Escola de Engenharia, Universidade Federal do Rio Grande do Sul, Porto Alegre. Disponível em: <https://www.lume.ufrgs.br/handle/10183/4711>. Acesso em: 11 ago. 2021.
- Liebel, M. J. & Fogliatto F. S. (2005). Método para previsão de receita tributária. In: Encontro Nacional de Engenharia de Produção, 25, 2005. Porto Alegre. *Anais...* Porto Alegre: ENEGEP, 2005. Disponível em: http://www.abepro.org.br/biblioteca/enegep2005_enegep0606_1542.pdf.
- Morettin, P. A. & TOLOI, C. M. C. (2004). *Análise de Séries Temporais*. São Paulo: Edgard Blucher.
- Oliveira, F. A. (2009). *Economia e política das finanças públicas no Brasil*. São Paulo: Hucitec.
- Shumway, R. H.; Stoffer, D. S. (2011) *Time Series Analysis and its Applications*. 3 ed. New York: Springer.
- Silva, P. M. & Weider, W. (2012) O planejamento não trata apenas das decisões sobre o futuro, mas questiona qual é o futuro das decisões que as organizações tomam. *Revista GUAL*, Florianópolis, v. 5, n. 4, p. 209-227, Edição Especial. 2012. Disponível em: <https://periodicos.ufsc.br/index.php/gual/article/view/1983-4535.2012v5n4p209/23688>. Acesso em: 11 ago. 2021.