

# Contribuição da modelagem de valores atípicos na previsão da arrecadação do ICMS do Estado de Minas Gerais

*Benefits of modeling outliers for forecasting state tax collection in Minas Gerais State*

Caio Peixoto Chain

Economista pela Universidade Federal Rural do Rio de Janeiro,  
Doutorando em Administração pela Universidade Federal de  
Lavras – UFLA.  
Lavras, MG [Brasil]  
caiochain@hotmail.com

Daniel Fonseca Costa

Contador, Doutorando em Administração pela Universidade  
Federal de Lavras – UFLA, Professor do Departamento de  
Gestão do Instituto Federal de Minas Gerais – IFMG/Campus  
Formiga.  
Formiga, MG [Brasil]

Naiara Leite dos Santos Sant'Ana

Doutoranda em Administração pela Universidade Federal de  
Minas Gerais – UFMG, Professora da Universidade Federal de  
Juiz de Fora – UFJF, Campus Governador Valadares.  
Governador Valadares, MG [Brasil]

Gideon Carvalho de Benedicto

Doutor em Contabilidade e Controladoria pela Universidade  
de São Paulo – USP, Professor do Programa de Pós-Graduação  
em Administração da Universidade Federal de Lavras – PPGA/  
UFLA.  
Lavras, MG [Brasil]

## Resumo

A previsão de arrecadação tributária é considerada uma ferramenta útil aos gestores públicos, além de ser uma obrigatoriedade pela Lei de Responsabilidade Fiscal. Assim, neste estudo, objetivou-se estimar um modelo preditivo da arrecadação de ICMS pelo governo de Minas Gerais para o período de janeiro de 1998 a agosto de 2011. O método utilizado foi a análise de séries temporais por meio dos modelos da família Box-Jenkins. Como resultado foi verificado que o método SARIMAX, ao considerar eventos diferentes do padrão histórico da série, apresentou melhor desempenho em relação às medidas de erro quando comparado aos métodos ARIMA, ARFIMA e SARIMA. Concluiu-se que a modelagem de valores atípicos contribuiu para uma melhor previsão das receitas de ICMS em Minas Gerais, assim, deveria ser levada em consideração pelos gestores públicos.

**Palavras-chave:** Finanças públicas. Gestão tributária. Gestores públicos. Modelos Box-Jenkins. Valores atípicos.

## Abstract

Tax revenue forecasting is considered a useful tool for policy makers and is also required by law (*Lei de Responsabilidade Fiscal*). This study aimed to estimate a predictive model for state tax (ICMS) collection by the government of Minas Gerais from January 1998 to August 2011. The method used was time series analysis through Box-Jenkins' family of models. According to the results, the SARIMAX method, which considered different events of the historical pattern of the series, showed better performance with regard to error measurements compared to the ARIMA, ARFIMA and SARIMA methods. We concluded that the modeling of outliers contributed to a better forecasting of state tax (ICMS) revenue in Minas Gerais and, consequently, should be taken into account by policy makers.

**Key words:** Box-Jenkins Models. Outliers. Public administrators. Public finance. Tax management.



## 1 Introdução

A estimativa sobre o futuro de um determinado fenômeno é de grande importância na tomada de decisão, seja qual for o ambiente em que ocorra a observação deste. Nas atividades econômicas e gerenciais, a previsão sobre o comportamento de uma variável serve de ferramenta estratégica para os agentes econômicos, viabilizando uma redução dos níveis de incerteza.

Desse modo, os modelos estatísticos de previsão emergem com a finalidade de buscarem, no comportamento histórico das séries temporais, respostas para os acontecimentos futuros (WALTER et al., 2013). Assim, os modelos de séries temporais oferecem técnicas de estimação e avaliação que produzem resultados de previsão em diferentes intervalos de tempo (CHEN; BLOOMFILELD; CUBBAGE, 2008).

Além disso, os modelos de previsões podem servir de apoio para a elaboração de orçamentos, permitindo a compreensão dos resultados futuros e auxiliando os gestores nas decisões orçamentárias (COSTA et al., 2014). Desta forma, esses modelos podem ser úteis para os Estados, auxiliando-os na melhoria da elaboração orçamentária e no planejamento da arrecadação tributária (SANTOS; LIMA, 2006).

Ao considerar que o Estado é um agente econômico que tem sua fonte de receita na arrecadação tributária, Castelar, Ferreira e Linhares (1996) ressaltaram que a previsão desta informação é crucial para o planejamento da máquina estatal, no que diz respeito ao custeio, investimentos e endividamento. No curto prazo, a previsão tem como propósitos auxiliar na gestão do fluxo de receita e atuar como indicador estratégico em relação às mudanças conjunturais. No longo prazo, as estimativas fornecem subsídios macroeconômicos para a adoção de medidas de políticas econômicas que viabilizem mudanças estruturais.

No Brasil, a necessidade de previsão com a arrecadação tributária federal, estadual e municipal também encontra respaldo jurídico no art. 12 da Lei Complementar 101 de 4/5/2000 (BRASIL, 2000), conhecida como Lei de Responsabilidade Fiscal. Essa lei estabelece a obrigatoriedade da previsão de receitas tributárias, por meio de métodos e premissas adequadas para os devidos cálculos, englobando as esferas de governo municipal, estadual e federal.

O sistema tributário brasileiro está dividido entre a União, os Estados, o Distrito Federal e os Municípios. Essa matriz é definida pela Constituição Federal (CF) de 1988, que, por meio dos artigos 153 a 155, atribui a esses entes a competência de instituir e cobrar os impostos (BRASIL, 1988). Em geral, os tributos incidem sobre patrimônio, renda, operações financeiras e consumo. Os tributos sobre consumo, também denominados de indiretos, são aqueles que incidem sobre a produção, sobre a circulação de mercadorias e sobre os serviços, sendo repassados para o preço final dos produtos e serviços (PEGAS, 2006).

Dentre os tributos indiretos, destaca-se o Imposto sobre Operações Relativas à Circulação de Mercadorias e sobre Prestações de Serviços de Transporte Interestadual, Intermunicipal e de Comunicação (ICMS), que foi objeto de análise do atual estudo, pela sua importância na receita dos Estados e do Distrito Federal.

O ICMS é previsto pelo inciso II, do Artigo 155 da Constituição Federal e instituído pela Lei Complementar 87, de 13 de setembro de 1996, que define os aspectos gerais de aplicação desse imposto (BRASIL, 1988; 1996). A Lei Complementar nº 87, conhecida como Lei Kandir, foi relevante para as finanças dos Estados, por ter sido instituída com o objetivo de estimular a competitividade internacional, por meio da não incidência do ICMS sobre a saída para o exterior de produtos

primários e semielaborados (LEITÃO; IRFFI; LINHARES, 2012).

O inciso I do §2º do art. 155 da Constituição Federal, concomitante com o art. 19 da Lei Complementar nº 87 diz que o ICMS é um imposto não cumulativo (BRASIL, 1988; 1996), o que significa que o valor devido de ICMS em cada operação relativa à circulação de mercadorias ou prestação de serviços de transporte interestadual, intermunicipal e de comunicação será compensado com o montante de ICMS pago nas operações anteriores. Assim, a não cumulatividade demonstra que a arrecadação mensal do ICMS vai sendo realizada, pelos Estados, à medida que as mercadorias vão circulando pela cadeia de suprimentos, até chegar ao consumidor final. Este fato ressalta a importância de prever a arrecadação tributária futura para fins de planejamento da aplicação do erário.

No entanto, a arrecadação de ICMS, assim como qualquer série temporal, pode ser afetada por eventos externos, tais como alterações na legislação, na política econômica, que tendem a gerar valores atípicos (ou *outliers*). Estes podem ser entendidos como observações que os modelos estatísticos não estão aptos a prever devido ao seu comportamento fora do padrão histórico (CHEN; LIU, 1993).

A previsão da arrecadação de ICMS de Minas Gerais (MG), utilizando modelos da família Box e Jenkins (1976), foi estudada por Santos e Lima (2006) e Pessoa, Coronel e Lima (2013). No primeiro caso foi incorporada a sazonalidade pelo método SARIMA (Seasonal Autoregressive Integrated Moving Average/Autorregressivo de Médias Móveis Sazonais) e, no segundo, os componentes relativos à memória de curto e longo prazo pelos métodos ARIMA (Autoregressive Integrated Moving Average /Autorregressivos integrados de Médias Móveis) e ARFIMA (Autoregressive Fractionally Integrated Moving-Average /Autorregressivos Fracionalmente

Integrados de Médias Móveis). No entanto, nestas análises, assume-se que o comportamento futuro é determinado pelo passado ao desconsiderar a presença de valores atípicos na série. Diante do exposto, este estudo tem a seguinte hipótese: a modelagem dos *outliers* na arrecadação do ICMS de Minas Gerais tende a reduzir consideravelmente o erro de previsão.

O objetivo neste trabalho foi estimar um modelo de previsão da arrecadação de ICMS de Minas Gerais para o período de janeiro de 1998 a agosto de 2011. Especificamente, pretendeu-se considerar os valores atípicos na previsão deste tributo, comparar as principais medidas de erro do modelo estimado com os modelos que desconsideram essa característica, assim como verificar sua validade para previsão de dados não amostrados 12 meses à frente.

## 2 Revisão teórica

Estudos sobre previsão de receitas tributárias estão amplamente difundidos na literatura nacional e internacional devido à importância do tema. Os diversos autores que abordam a temática buscam, em geral, ajustar diferentes modelos que resultem em um melhor poder de previsão, ou seja, um nível aceitável entre o valor real e o estimado.

Uma exaustiva revisão de literatura em nível internacional sobre previsão fiscal pode ser encontrada em Leal et al. (2008). Segundo os autores, em especial para o caso da União Europeia, a crescente relevância do tema está relacionada ao Tratado de Maastricht e ao Pacto de Estabilidade e Crescimento que elevaram a vigilância orçamentária multilateral. Os maiores desafios apontados foram a mensuração do desempenho das previsões (viés, eficiência, acurácia e causas dos erros de previsão) e o processo decisório para uma melhor previsão (método, dados, horizonte e nível de



desagregação). Por exemplo, Favero e Marcellino (2005) utilizaram diversos tipos de modelagens estatísticas como modelos Autorregressivos de Média Móveis (ARIMA), Vetores Autorregressivos (VAR), Equações Simultâneas (ES) e a Forecast Pooling (FP) para previsão de variáveis fiscais e macroeconômicas na Zona do Euro, em 2002, com base em dados semestrais de 1981 a 2001. Utilizando como medida a raiz do Erro Quadrado Médio de Previsão (REMP) um passo à frente da série de receita fiscal, o modelo ARIMA apresentou os melhores resultados para França e Itália, enquanto para Alemanha e Espanha foram a FP e a ES, respectivamente.

Além dos métodos estatísticos tradicionais, os métodos computacionais também apresentam crescente contribuição na modelagem e previsão de receitas tributárias. Li-xia, Yi-qi e Liu (2011) propuseram a combinação dos métodos Support Vector Machine com o Particle Swarm Optimization para previsão de tributos na China, no período anual de 1990 a 1997. O resultado do método proposto foi superior ao das redes neurais. O Erro Percentual Médio de Previsão (EPMP) fora da amostra foi 1,18% pelo primeiro método, e 5,20% pelo segundo.

A previsão de tributos no Brasil encontra respaldo na literatura recente. Em geral, a maior fonte de arrecadação tributária dos Estados brasileiros é o ICMS. No caso de Minas Gerais, este tributo foi responsável por cerca de 80% da receita tributária do Estado durante o período de 1998 a 2011, segundo a Secretaria de Estado de Fazenda de Minas Gerais (SEF-MG, 2012).

Contreras e Cribari-Neto (2006) buscaram estimar a arrecadação de ICMS de Pernambuco (PE), Rio de Janeiro (RJ) e São Paulo (SP) para o período mensal de 1994 a 2004, com correção de *outliers*. Os autores compararam métodos estatísticos, computacionais e a combinação de ambos para fazerem previsões para 2005. Considerando

seis passos à frente em 2004, o menor EPMP para PE e RJ foi 5,60% e 6,04%, respectivamente, por meio de redes neurais. Para São Paulo, o menor EPMP foi 1,69% pelo método SARIMA.

O ICMS de Santa Catarina foi previsto por Corvalão, Samohyl e Brasil (2010), os quais utilizaram regressões dinâmicas em dados mensais de 1995 até 2001. Este método, além da variável arrecadação de ICMS, também levou em consideração: atividade industrial; consumo de energia elétrica, combustíveis e cimento; e a quantidade de consultas das empresas ao Serviço de Proteção ao Crédito. Os valores observados de 2001 foram comparados com a previsão da regressão dinâmica e com o modelo vigente no Departamento de Orçamento, ARIMA (1,0,2) e computaram um EPMP de 2,51% e 4,63%, respectivamente. Para os quatro meses de 2002 utilizados como estimativas fora da amostra o EPMP da regressão foi 2,55%.

A metodologia Box-Jenkins foi utilizada por Clemente e Clemente (2011) para estimar as receitas de ICMS de Santa Catarina com base em dados mensais de 2001 a 2011. Devido à identificação da sazonalidade de 12 meses, estes autores trabalharam com a comparação entre modelos SARIMA e elegeram como melhor opção um SARIMA (2,1,0) (0,1,1). A previsão fora da amostra foi realizada de 2011 a 2012, logo não foram reportadas as medidas de erro.

Conforme mencionado na introdução, em Minas Gerais a previsão do ICMS foi estudada por Santos e Lima (2006) e por Pessoa, Coronel e Lima (2013). Os autores Santos e Lima (2006) utilizaram uma série de arrecadações de janeiro de 1992 a maio de 2004 e concluíram que, dentre os modelos estimados, o melhor ajuste foi um SARIMA (12,1,12)(0,1,1), sendo utilizado para previsões de junho de 2004 a maio de 2005, e apresentou um EPMP de 4,56%.

Pessoa, Coronel e Lima (2013) utilizaram em seu estudo o período mensal de janeiro de 1998 a agosto de 2011. Os modelos com melhor ajuste e utilizados para comparação foram um ARIMA (1,0,1) e um ARFIMA (1,0.36,1) que apresentaram um EPMP dentro da amostra de 6,13% e 6,03%, respectivamente. É válido ressaltar que outras medidas de erro, como REMP e Erro Absoluto Médio (EAP), apuradas pelos autores, indicaram o modelo ARIMA para previsões mais acuradas.

### 3 Materiais e métodos

#### 3.1 Modelo analítico

Uma série temporal univariada é composta pela soma de três componentes, ou seja,  $Y_t = T_t + S_t + a_t$  em que  $T_t$  é a tendência,  $S_t$  a sazonalidade e  $a_t$  o termo aleatório (ou erro). Em geral, supõe-se que  $a_t$  seja puramente aleatório ou represente um ruído branco independente de média zero e variância constante (SÁFADI, 2004).

Como caso geral dos modelos da família Box e Jenkins (1976), um modelo Autorregressivo de Médias Móveis (ARIMA) é útil para descrever séries de tempo que não possuem uma média constante ao longo do tempo (não estacionárias) e que demandam poucos parâmetros de ajuste (parcimônia). Um modelo ARIMA que leve em consideração uma tendência ao longo do tempo, apresenta ordem (p,d,q) e pode ser formalizado pela equação:

$$\phi(B)(1 - B)^d Y_t = \theta(B)a_t \quad (1)$$

Em que:

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  o polinômio autorregressivo de ordem  $p$ ;

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$  o polinômio de médias móveis de ordem  $q$ ;

$B$  o operador de defasagem, em que  $B^d Y_t = Y_{t-d}$  e  $d$  é a ordem da diferença necessária para tornar a série estacionária retirando sua tendência.

Quando o conjunto de dados em análise apresenta um comportamento periódico dentro de um ano (sazonalidade), pode-se considerar que seu modelo possui uma componente sazonal. O modelo sazonal SARIMA (p,d,q)(P,D,Q)<sub>s</sub> pode ser descrito como:

$$\phi_p(B)\Phi_p(B^s)(1 - B)^d(1 - B)^D Y_t = \theta_q(B)\Theta_Q(B^s)$$

Em que:  $B)\Phi_p$  e  $(B)\Theta_Q$  são os polinômios de sazonalidade autorregressivos e de médias móveis de ordem  $P$  e  $Q$ , respectivamente, e  $D$  é o número de diferenças para eliminar a sazonalidade.

O operador geral de diferença, tanto para tendência quanto para sazonalidade, pode ser definido quando duas observações que estão distantes entre si por  $s$  intervalos de tempo apresentam comportamento similar. Esse operador ( $\nabla_s$ ) pode ser formalizado por:  $\nabla_s Y_t = (1-B)^s Y_t = Y_t - Y_{t-s}$ .

Segundo Kaiser e Maravall (2001), uma série temporal com *outliers* pode ser representada por:

$$Y_t^* = \sum_{j=1}^k \xi_j(B)\omega_j I_t^{(\tau_j)} + Y_t$$

Sendo  $Y_t^*$  a série observada (com a presença de *outliers*);  $Y_t$  o processo ARIMA;  $\omega_j$  é o impacto inicial do *outlier* no tempo  $t = \tau_j$ ;  $I_t^{(\tau_j)}$  é uma variável indicadora que assume o valor 1, quando  $t = \tau_j$ , e 0 caso contrário; e  $\xi_j(B)$  determina a dinâmica dos *outliers* conforme as seguintes possibilidades:

- *Additive Outlier* (AO):  $\xi_j(B) = 1$ ;
- *Level Shift* (LS):  $\xi_j(B) = 1/(1-B)$ ;
- *Temporary Change* (TC):  $\xi_j(B) = 1/(1-\delta B)$ ,  $0 < \delta < 1$ ;
- *Innovational Outlier* (IO):  $\xi_j(B) = \theta(B)/\phi(B)$ .



De acordo com Chen e Liu (1993), exceto pelo IO, o efeito destes *outliers* é independente do modelo estimado e AO e LS são casos específicos de TC, quando  $\delta=0$  e  $\delta=1$ , respectivamente. O efeito de AO é abrupto e temporário, o de LS é abrupto e permanente. Por sua vez, TC apresenta um impacto gradual e temporário a uma taxa de decaimento de “ $\delta$ ”. Já o IO pode causar diversos efeitos visto que depende do modelo ajustado para  $Y_t$ . Em geral esse efeito é explosivo e aumenta à medida que a ocorrência do fenômeno fica mais distante.

Em resumo, o método ARIMA, ao incorporar a sazonalidade e valores atípicos (uma ou mais variáveis exógenas binárias, no caso desse estudo  $I_t^{(\tau)}$ ) pode ser chamado de SARIMAX.

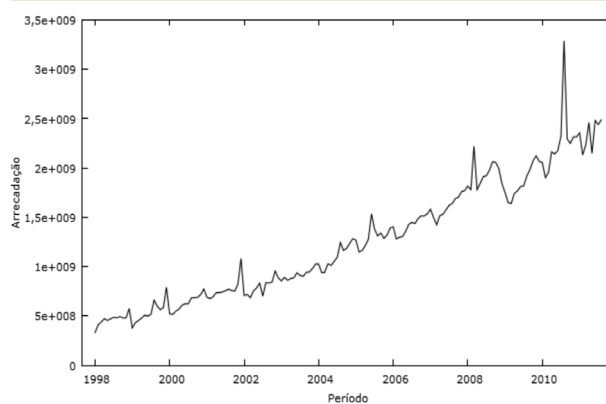
### 3.2 Fonte de dados

Para a composição da série temporal analisada neste estudo, foram coletados os dados com a arrecadação de ICMS de Minas Gerais, em reais, disponibilizados pela Secretaria Estadual de Fazenda (SEF-MG, 2012). Esta série foi referente ao período mensal de janeiro de 1998 a agosto 2011 e totalizou 164 observações.

## 4 Resultados e discussão

Na Figura 1, que representa a arrecadação de ICMS de Minas Gerais, foi possível verificar que a série possui uma tendência crescente e uma sazonalidade com picos no mês de dezembro. Essa sazonalidade é comum em dados que indicam níveis de atividade econômica, como o PIB, consumo de energia, nível de empregos. No entanto, foram aplicados os devidos testes estatísticos para confirmar a presença desses elementos.

Para verificar se a série é ou não estacionária, foi aplicado o teste de Dickey-Fuller Aumentado (ADF) proposto por Dickey e Fuller (1981). O ADF não rejeitou a hipótese nula de presença de



**Figura 1: Arrecadação de ICMS mensal de Minas Gerais, 1998-2011**

Fonte: Secretaria de Estado da Fazenda de Minas Gerais (SEF-MG, 2012).

raiz unitária, assim a série não foi considerada estacionária, pois foi confirmada a presença de uma tendência temporal. Para eliminar essa tendência, foi utilizada a diferenciação de primeira ordem da série original. Para corroborar os resultados foram aplicados os testes de Dickey-Fuller Aumentado (ADF) com constante e com tendência.

Para a hipótese de ausência de sazonalidade, o teste utilizado foi o de Kruskal-Wallis (DAGUM, 1988). A presença de um movimento periódico ao longo do tempo foi estatisticamente significativa na série original e na série em primeira diferença, logo a sazonalidade foi confirmada. Assim, para eliminar esse efeito, foi adotado o procedimento de diferenciação sazonal da série em primeira diferença.

Após a retirada da tendência e da sazonalidade, a série foi considerada estacionária, conforme os resultados dos testes que foram demonstrados na Tabela 1.

Em seguida, foi gerado o correlograma da série transformada, que pode ser visualizado na Figura 2. Assim, foi possível verificar os *lags* em que ainda existiam correlações significativas diferentes de zero, de maneira que estas pudessem ser modeladas por componentes autorregressivos (AR), de médias móveis (MA) e sazonais (SAR de SMA).

**Tabela 1: Resultado dos testes ADF e Kruskal-Wallis**

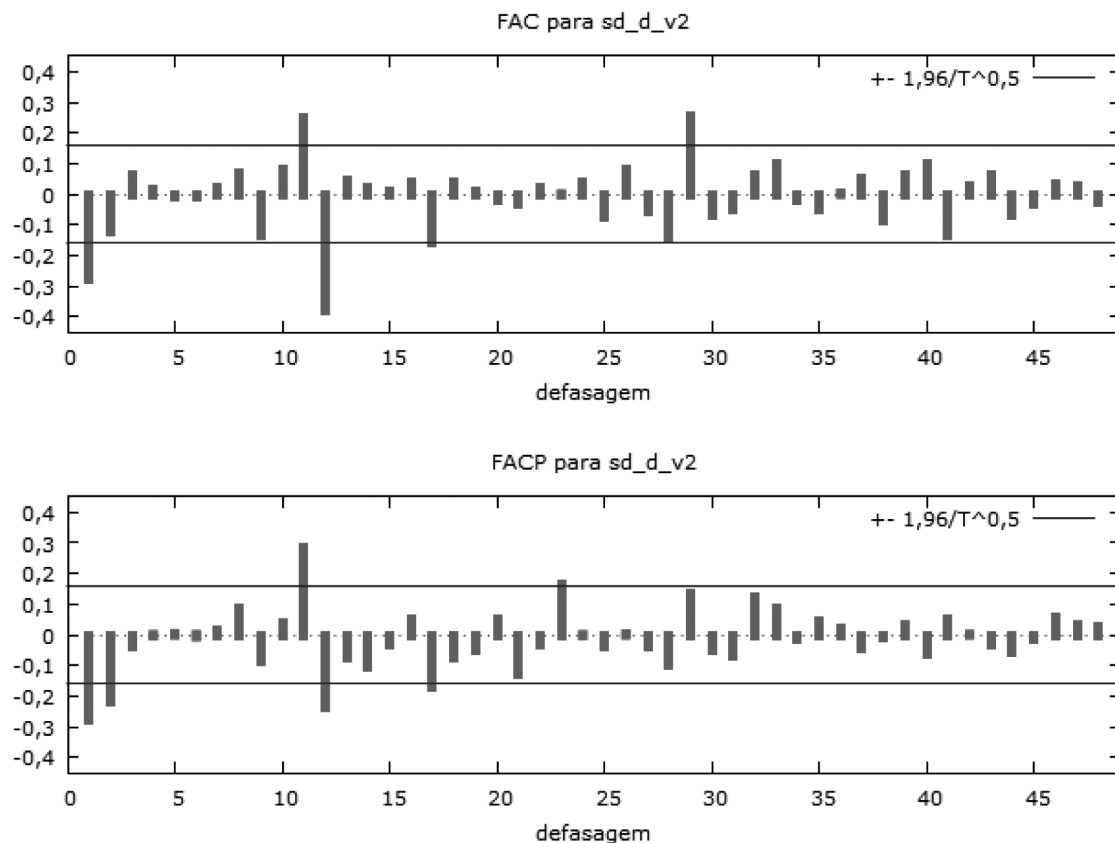
Teste	Hipótese nula	p-valor (série em nível)	p-valor (1ª diferença)	p-valor (dif. sazonal)
ADF com const.	Presença de raiz unitária	0,99	<0,001	<0,001
ADF com const. e tend.	Presença de raiz unitária	0,49	<0,001	<0,001
Kruskal-Wallis	Ausência de sazonalidade	<0,001	<0,001	0,97

Fonte: Os autores.

Pela Função de Autocorrelação Total pode-se verificar que somente o primeiro *lag* foi significativo, assim como a parte sazonal apresentou um *lag* significativo na 12ª defasagem. Esta inspeção foi utilizada para modelagem dos processos MA e SMA. Para a análise dos processos AR e SAR, foram considerados, respectivamente, os dois primeiros *lags* significativos, assim como a 12ª defasagem da Função de Autocorrelação Parcial.

Após a inspeção destas duas funções, diversos modelos foram ajustados. Segundo Walter et al. (2013), a metodologia Box-Jenkins é composta por três etapas cíclicas: (i) identificar o modelo por meio do correlograma, (ii) estimar os parâmetros AR e/ou MA e (iii) diagnosticar os resíduos. Os critérios de seleção de modelos utilizados foram o de Schwarz (1978) e de Akaike (1998), em conjunto com o melhor desempenho relativo às métricas de erros de previsão (EPMP, REQM e EAM).

O modelo selecionado foi um SARIMA (1,1,1) (1,1,1)<sub>12</sub>. Como podem ser visualizados na Tabela 2, os coeficientes estimados por este modelo foram estatisticamente significativos no nível de 1%, exceto pelo autorregressivo sazonal, significativo em 5%. Assim, foi demonstrado que a série de arrecadação ICMS de Minas Gerais é bem explicada por seus valores passados.

**Figura 2: Função de autocorrelação total (FAC) e parcial (FACP) da série transformada**

Fonte: Adaptado pelos autores com base no software Gretl.

Como diagnóstico do modelo, o teste de autocorrelação residual de Ljung e Box (1978) não rejeitou a hipótese nula (p-valor: 0,24), o que indicou que os resíduos não são correlacionados. O teste ARCH (ENGLE, 1982) não rejeitou a hipótese de que os resíduos apresentaram comportamento homoscedástico (p-valor: 0,99). Pelo teste de Jarque e Bera (1980), a hipótese de normalidade dos resíduos não foi aceita (p-valor: < 0,001).

**Tabela 2: Modelo SARIMA (1,1,1) (1,1,1) para arrecadação de ICMS de Minas Gerais**

	Coeficiente	Erro-padrão	Z	p-valor
Const.	1,03138e+06	155985	6,6121	<0,00001
phi_1	0,322414	0,0753006	4,2817	0,00002
Phi_1	-0,252035	0,125402	-2,0098	0,04445
theta_1	-0,999988	0,026129	-38,2712	<0,00001
Theta_1	-0,651647	0,117398	-5,5508	<0,00001

Critério de Schwarz = 6100,481  
 Critério de Akaike = 6082,377

Fonte: Adaptado pelos autores com base no *software* Gretl.

Com base no método proposto por Findley et al. (1998) e Gomez e Marvall (2001), os *outliers* encontrados foram tratados como do tipo aditivo (Additive Outlier – AO). Em outras palavras, o processo iterativo de controle de valores atípicos se deu com base na inclusão de uma variável *dummy* unitária em cada um dos resíduos que, em geral, excedeu 2,5 vezes o erro-padrão das estatísticas de avaliação da previsão.

Segundo Kaiser e Maravall (2001), o principal objetivo da correção de *outliers* é transformar os dados até que a hipótese de normalidade dos resíduos seja aceita. Foram tratados cinco *outliers* ao longo do período analisado até que o critério de normalidade dos resíduos fosse alcançado, a saber: 12/2001, 06/2005, 03/2008, 08/2010 e 05/2011. Foi então estimado um modelo SARIMAX (1,1,1) (1,1,1), apresentado na Tabela 3.

Não foram encontrados na literatura os fatores que ocasionaram estes *outliers*. No entanto, deve-se ressaltar que o objetivo do trabalho foi uma modelagem preditiva e não explicativa,

**Tabela 3: Modelo SARIMAX (1,1,1) (1,1,1) para arrecadação de ICMS de Minas Gerais**

	Coeficiente	Erro-padrão	Z	p-valor
Const.	1,02696e+06	234859	4,3727	0,00001
phi_1	0,752274	0,0614338	12,2453	<0,00001
Phi_1	-0,0117554	0,115157	-0,1021	0,91869
theta_1	-1	0,0350624	-28,5206	<0,00001
Theta_1	-0,548104	0,0856778	-6,3973	<0,00001
12/2001	2,47002e+08	3,93449e+07	6,2779	<0,00001
06/2005	1,90881e+08	3,92256e+07	4,8662	<0,00001
03/2008	4,98028e+08	3,93448e+07	12,6581	<0,00001
08/2010	9,80543e+08	4,19613e+07	23,3678	<0,00001
05/2011	-3,01172e+08	4,44848e+07	-6,7702	<0,00001

Critério de Schwarz = 5878,597  
 Critério de Akaike = 5845,407

Fonte: Adaptado pelos autores com base no *software* Gretl.

logo os resultados não foram prejudicados. A própria definição de *outlier* proposta por Chen e Liu (1993) considera que a estimação deste fenômeno se aplica quando não existem informações *a priori* sobre a causa ou o tempo em que o este foi gerado.

Todos os valores atípicos incluídos no modelo foram significativos a 1%, ou seja, foi demonstrado que de fato influenciaram a arrecadação de ICMS de MG, conforme demonstrado na Tabela 3. Vale destacar que exceto no período de 05/2011, todos os demais representaram momentos de impacto positivo na arrecadação. O de maior relevância foi em agosto de 2010, quando foi arrecadado, aproximadamente, R\$1bilhão acima do que era esperado pelo padrão histórico da série.

Após o controle dos *outliers*, os resíduos do modelo SARIMAX (1,1,1)(1,1,1) mantiveram as condições de ausência de autocorrelação serial (p-valor: 0,53) e homoscedasticidade (p-valor: 0,05). Adicionalmente, a normalidade dos resíduos não foi rejeitada no nível de significância de 10% (p-valor: 0,13), ou seja, os resíduos apresentaram comportamento do tipo “ruído branco”. Os critérios de Schwarz e Akaike também foram reduzidos consideravelmente, indicando melhor ajuste do modelo.

Para previsões da arrecadação de ICMS dentro da amostra – janeiro de 1998 a agosto de 2011



– os modelos com sazonalidade apresentaram melhor desempenho em relação às principais medidas de erro. Em especial, o modelo SARIMAX (1,1,1) (1,1,1) que reduziu, aproximadamente, pela metade suas medidas de erro em relação aos modelos ARIMA (1,0,1) e ARFIMA (1,0.36,1), conforme Tabela 4. A medida que sofreu maior impacto com a inclusão dos *outliers* foi a REQM que decresceu de uma ordem superior aos R\$100 milhões para o valor de R\$54.530.725,28.

**Tabela 4: Comparação entre medidas de erro dos modelos ajustados**

	ARIMA (1,0,1)	ARFIMA (1,0.36,1)	SARIMA (1,1,1)(1,1,1)	SARIMAX (1,1,1)(1,1,1)
EPMP	6,13%	6,03%	4,78%	3,52%
REQM	125.995.275,00	130.010.398,88	121.860.576,07	54.530.725,28
EAM	71.121.694,83	73.842.444,07	67.934.000,00	41.795.000,00

Fonte: Pessoa, Coronel e Lima (2013) e adaptado pelos autores com base no *software* Gretl.

Este conjunto de informações confirmou a hipótese levantada por esta pesquisa de que a inclusão de valores atípicos no modelo de previsão de ICMS de Minas Gerais foi significativa para reduzir as diferenças entre os valores reais e os estimados.

O modelo SARIMAX (1,1,1)(1,1,1), considerado de melhor desempenho, foi utilizado para a previsão fora da amostra (09/2011 a 08/2012). O EPMP foi computado para os 12 meses, visando a verificar a validade do controle de *outliers* em estimativas para dados não amostrados. Estas previsões foram comparadas com os valores reais disponibilizados pela SEF-MG e foram apresentadas na Tabela 5.

Assim como nas previsões dentro da amostra, o modelo SARIMAX (1,1,1)

(1,1,1) se mostrou eficiente para realizar estimativas fora da amostra, com um EPMP de 2,72%. Deve-se ressaltar que este tipo de estimativa fora da amostra é que, de fato, representa interesse para os gestores públicos em seu processo de tomada de decisão. Exceto no mês de dezembro, todos os valores reais ficaram dentro das margens de erro que foram calculadas com um intervalo de confiança de 95%.

No entanto, a modelagem de séries temporais deve ser revisada periodicamente, pois à medida que cada novo valor é realizado um processo (S) ARMA distinto pode ser mais adequado, assim como algum *outlier* pode ter seu efeito diluído enquanto outros se tornam significativos. Conforme apresentado por Willoughby e Guo (2008), 48% dos governos estaduais dos Estados Unidos atualizam seus métodos de previsão de receitas periodicamente (em um ou quatro meses), 45% de acordo com a demanda e os demais não possuem revisão formal.

## 5 Conclusões

O objetivo neste trabalho foi modelar a arrecadação de ICMS de Minas Gerais, consideran-

**Tabela 5: Arrecadação de ICMS/MG real, estimada, margens de erro (R\$) e erro (%), 09/11 a 08/12**

	Real	Estimado	Margem inferior	Margem superior	Erro percentual
Set./11	2.589.989.989,65	2.518.330.000,00	2.409.290.000,00	2.627.380.000,00	0,23%
Out./11	2.549.456.054,60	2.522.920.000,00	2.386.460.000,00	2.659.370.000,00	0,09%
Nov./11	2.545.422.921,15	2.564.780.000,00	2.415.020.000,00	2.714.540.000,00	0,06%
Dez./11	2.788.468.177,37	2.535.640.000,00	2.378.850.000,00	2.692.430.000,00	0,76%
Jan./12	2.663.415.267,52	2.546.070.000,00	2.385.440.000,00	2.706.700.000,00	0,37%
Fev./12	2.286.807.937,10	2.385.380.000,00	2.222.620.000,00	2.548.150.000,00	0,36%
Mar./12	2.519.049.903,16	2.437.610.000,00	2.273.650.000,00	2.601.570.000,00	0,27%
Abr./12	2.673.853.148,48	2.616.830.000,00	2.452.200.000,00	2.781.460.000,00	0,18%
Mai./12	2.525.834.434,30	2.619.180.000,00	2.454.170.000,00	2.784.200.000,00	0,31%
Jun./12	2.677.055.992,91	2.660.950.000,00	2.495.720.000,00	2.826.180.000,00	0,05%
Jul./12	2.678.343.067,40	2.686.270.000,00	2.520.920.000,00	2.851.620.000,00	0,02%
Ago./12	2.716.834.194,49	2.725.000.000,00	2.559.580.000,00	2.890.410.000,00	0,03%
					EPMP
					2,72%

Fonte: SEF-MG e adaptado pelos autores com base no *software* Gretl.

do a ocorrência de valores atípicos. Os modelos ARIMA e ARFIMA contabilizaram um EPMP de 6,13% e 6,03%. Ao se considerar a sazonalidade, o melhor ajuste foi um SARIMA (1,1,1)(1,1,1) com o EPMP de 4,78%. Com a modelagem de cinco valores atípicos presentes nesta série foi obtido um modelo SARIMAX (1,1,1)(1,1,1) com EPMP de 3,52%. Este último modelo foi utilizado para previsão de valores fora da amostra e gerou um EPMP de apenas 2,72%.

Assim, pode-se constatar que o atual trabalho contribuiu para corroborar a hipótese de que o tratamento dos *outliers* tenderia a reduzir o erro na previsão da arrecadação de ICMS de Minas Gerais. Na prática, isso implica dizer que os gestores públicos que não incorporarem os valores atípicos em suas estimativas, provavelmente, estarão trabalhando com valores mais distantes da realidade, prejudicando o processo de tomada de decisões.

Foi considerada como limitação do estudo a impossibilidade de identificação das causas geradoras dos valores atípicos na literatura. O conhecimento dos gestores da SEF-MG sobre estes períodos específicos pode ser útil para incorporar este tipo de fenômeno nas atualizações do modelo ao longo dos próximos anos.

Para estudos futuros recomenda-se a combinação de métodos de previsão aplicados à arrecadação de ICMS de MG, objetivando a redução do erro de previsão, principalmente para dados não amostrados. Técnicas promissoras como a combinação linear do método estatístico SARIMAX com métodos computacionais (Redes Neurais, por exemplo) ainda podem ser testadas.

## Referências

AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In: PARZEN, E.; TANABE, K.; KITAGAWA, G (Ed.). *Selected papers of Hirotugu Akaike*. Springer: New York, 1998. p. 199-213.

BOX, G. E. P.; JENKINS, G. M. *Time series analysis, forecasting and control*. San Francisco: Holden-Day, 1976.

BRASIL. Constituição da República Federativa do Brasil de 1988. Brasília, DF: Senado, 1988.

BRASIL. Lei Complementar n. 87, de 13 de setembro de 1996. Dispõe sobre o imposto dos Estados e do Distrito Federal sobre operações relativas à circulação de mercadorias e sobre prestações de serviços de transporte interestadual e intermunicipal e de comunicação, e dá outras providências. *Diário Oficial da República Federativa do Brasil*, Brasília, 16 set. 1996; Seção 1, p. 18261.

BRASIL. Lei Complementar n. 101, de 4 de maio de 2000. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. *Diário Oficial da República Federativa do Brasil*, Brasília, DF, 5 maio. 2000.

CASTELAR, I.; FERREIRA, R. T.; LINHARES, F. C. Modelos de previsão para o ICMS do Estado do Ceará. *Revista Econômica do Nordeste*, Fortaleza, v. 27, n. 4, p. 583-606, 1996.

CHEN, C.; LIU, L. Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, v. 88, n. 421, p. 284-297, 1993.

CHEN, R. J. C.; BLOOMFIELD, P.; CUBBAGE, F. W. Comparing forecasting models in tourism. *Journal of Hospitality & Tourism Research*, v. 32, n. 1, p. 3-21, 2007.

CLEMENTE, A.; CLEMENTE, L. T. Aplicação da metodologia Box-Jenkins para previsão do ICMS do estado do Paraná de agosto de 2011 a julho de 2012. *Economia & Tecnologia*, Ano 7, v. 27, 2011.

CONTRERAS, J. C. S.; CRIBARI-NETO, F. Previsão da arrecadação do ICMS no Brasil através de Redes Neurais. *Revista Brasileira de Estatística*, Rio de Janeiro, v. 67, n. 227, p. 7-43, jul./dez., 2006.

CORVALÃO, E. D.; SAMORYL, R. W.; BRASIL, G. H. Forecasting the collection of the state value added tax (ICMS) in Santa Catarina: the general to specific approach in regression analysis. *Brazilian Journal of Operations & Production Management*, v. 7, n. 1, p. 105-121, 2010.

COSTA, D. F. et al. Estimando o orçamento em uma empresa varejista: uma abordagem por meio de vetor autorregressivo – VAR. *Exacta – EP*, São Paulo, v. 12, n. 3, p. 279-291, 2014.

DAGUM, E. B. The X-11-ARIMA/88 seasonal adjustment method: foundations and user's manual. Ottawa: Statistics Canada, 1988.

DICKEY, D. A.; FULLER, W. A. Likelihood ratio statistics for auto-regressive time series with a unit root. *Econometrica*, v. 49, n. 4, p. 1057-1072, 1981.

- ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, p. 987-1007, 1982.
- FAVERO, C. A.; MARCELLINO, M. Modelling and forecasting fiscal variables for the euro area. *Oxford Bulletin of Economics and Statistics*, v. 67, n. s1, p. 755-783, 2005.
- FINDLEY, D. F. et al. New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. *Journal of Business & Economic Statistics*, v. 16, n. 2, p. 127-152, 1998.
- GOMEZ, V.; MARAVALL, A. Automatic modeling methods for univariate series. In: PEÑA, D.; TIAO, G. C.; TSAY, R. S. (Ed.). *A course in time series analysis*. Hoboken: Wiley & Sons, 2001, p. 171-201.
- JARQUE, C. M.; BERA, A. K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, v. 6, n. 3, p. 255-259, 1980.
- KAISER, R.; MARAVALL, A. Seasonal outliers in time series. *Estadística: Journal of the Interamerican Statistical Institute*, v. 53, n. 160/161 (Special Issue on Time Series Analysis), p. 97-142, 2001.
- LEAL, T. et al. Fiscal Forecasting: Lessons from the Literature and Challenges. *Fiscal Studies*, v. 29, n. 3, p. 347-386, 2008.
- LJUNG, G. M.; BOX, G. E. P. On measure of lack of fit in time series models. *Biometrika*, v. 65, n. 2 p. 297-303, 1978.
- LEITÃO, A. L.; IRFFI, G.; LINHARES, F. Avaliação dos efeitos da Lei Kandir sobre a arrecadação de ICMS no Estado do Ceará. *Planejamento e Políticas Públicas*, v. 39, n. 2, p. 37-63, 2012.
- LI-XIA, L.; YI-QI, Z.; LIU, X. Tax forecasting theory and model based on SVM optimized by PSO. *Expert Systems with Applications*, v. 38, n. 1, p. 116-120, 2011.
- PÊGAS, P. H. *Manual de contabilidade tributária*. 4. ed. Rio de Janeiro: Freitas Bastos, 2006.
- PESSOA, F. M. C.; CORONEL, D. A.; LIMA, J. E. Previsão de arrecadação de ICMS para o estado de Minas Gerais: uma comparação entre modelos Arima e Arfima. *Revista Brasileira de Gestão e Desenvolvimento Regional*, v. 9, n. 2, p. 47-64, 2013.
- SÁFADI, T. Uso de séries temporais na análise de vazão de água na represa de Furnas. *Ciência e Agrotecnologia*, v. 28, n. 1, p. 142-148, 2004.
- SANTOS, C. M. dos; LIMA, J. E. Análise de previsões da arrecadação do ICMS no Estado de Minas Gerais. *Revista de Economia e Administração*. v. 5, n. 4, p. 413-423, 2006.
- SCHWARZ, G. Estimating the dimension of a model. *The Annals of Statistics*, v. 6, n. 2, p. 461-464, 1978.
- SEF-MG. Secretaria de Estado de Fazenda de Minas Gerais. *Evolução da receita estadual*. Belo Horizonte: SEF, 2012. Disponível em: <<http://www.fazenda.mg.gov.br/>>. Acesso em: 17 dez. 2014.
- WALTER, O. M. F. C. et al. Aplicação de um modelo SARIMA na previsão de vendas de motocicletas. *Exacta – EP*, São Paulo, v. 11, n. 1, p. 77-88, 2013.
- WILLOUGHBY, K. G.; GUO, H. The state of the art: Revenue forecasting in US state governments. In: SUN, J.; LYNCH, T. D. (Ed.). *Government budget forecasting: theory and practice*. Boca Raton: Taylor & Francis Group, 2008, p. 27-42.

Recebido em 26 maio 2015 / aprovado em 30 set. 2015

**Para referenciar este texto**

CHAIN, C. P. et al. Contribuição da modelagem de valores atípicos na previsão da arrecadação do ICMS do Estado de Minas Gerais. *Exacta – EP*, São Paulo, v. 13, n. 2, p. 239-249, 2015.

