

Modelação e previsão das receitas do estado do Imposto sobre o Valor Acrescentado.

BRUNO NASCIMENTO*

Lisboa School of Economics & Management
bmsnascimento@gmail.com

Resumo

O presente trabalho motivado pela importância atual da previsão da receitas governamentais, mais concretamente das obtidas através do IVA, procura dentro da previsão univariada encontrar modelos de previsão competentes. Ao longo deste trabalho foram usados modelos de alisamento exponencial e modelos ARIMA. Esta procura de preditores competentes estendeu-se a catorze modelos e duas combinações entres estes.

Os resultados apontam para uma boa precisão dos modelos de Holt-Winters a 1 passo e a 12 passos. No caso das combinações das previsões, estas não aparentam melhorar a precisão das previsões.

Palavras-chave: IVA, sazonalidade, outliers, Holt-Winters, SARIMA e combinação de previsões.

I. INTRODUÇÃO.

No presente trabalho pretende-se modelar e prever as receitas dos impostos do Estado português, mais concretamente a receitas referentes ao Imposto sobre o Valor Acrescentado (IVA).

Segundo o Preâmbulo do código do IVA, este visa tributar todo o consumo em bens materiais e serviços, abrangendo na sua incidência todas as fases do circuito económico, desde a produção ao retalho, sendo, porém, a base tributável limitada ao valor acrescentado em cada fase.

O pagamento do imposto é feito em duas modalidades. Segundo o ponto número 1 do artigo 41º, do código do IVA, as empresas com um volume de negócios inferior a 650 000 € têm até ao dia 15 do 2º mês seguinte ao trimestre do ano civil a que respeitam o imposto, enquanto as empresas com um volume de negócios igual ou superior a 650 000 € têm até ao dia 10 do 2º mês seguinte àquele a que respeitam as operações. Contudo no ponto número 2, do mesmo artigo, é dada a possibilidade às empresas com apuramento do IVA trimestral (com volume de negócios inferior a 650 000 €) a passarem para a modalidade mensal, isto é, as empresas com volume de negócios inferior a 650 000 € podem optar entre o regime de apuramento trimestral e pagamento até ao dia 15 do 2º mês seguinte, ou o regime de apuramento mensal com o pagamento a ser efetuado até ao dia 10 do segundo mês seguinte ao mês apurado.

A previsão das receitas mensais do IVA é impor-

tante, no sentido em que esta componente é bastante relevante na receitas correntes. As receitas do IVA correspondem a cerca de 33% das receitas correntes do Estado português. Numa ótica de gestão das finanças públicas, uma ideia antecipada das receitas, a concretizar no futuro, é útil para a planificação do orçamento do Estado. Outra visão possível é em termo da política orçamental, a aplicação do imposto está inerente a três taxas de imposto (tanto as taxas como os respetivos bens abrangidos são definidas pelo Governo), como tal com o objetivo de aumentar ou baixar as receitas, dadas as previsões para o período, o governo pode alterá-las.

Os dados usados neste trabalho proveem da base de dado do Banco de Portugal, são de carácter mensal e os dados representam o período de 2000 a 2013. Esta amostra será dividida em três:

- De janeiro 2000 a dezembro 2011: para a modelação da série;
- De janeiro 2012 a dezembro 2012: para testar as capacidades de cada modelo no campo da previsão;
- De janeiro 2013 a dezembro 2013: para comparar a qualidade dos modelos testados no grupo anterior com combinações dos modelos.

No ponto número dois do trabalho, será feita uma caracterização da série onde procurar-se-á decompor a série nas suas principais componentes, estacionarizar a serie (caso esta não o seja) e detetar os

*Com o seguinte número de processo 40778.

seus *outliers*.

No ponto número três, será exposta toda a metodologia a aplicar no trabalho, entre os quais, o modelo de alisamento exponencial escolhido, os modelos ARIMA e a forma de combinar as previsões. Num quarto ponto, serão apresentados só resultados da aplicação da metodologia expressa no ponto anterior. Dar-se-á grande ênfase na qualidade de previsão de cada modelo. No quinto e ultimo ponto, serão enunciadas as principais conclusões do trabalho.

II. CARACTERIZAÇÃO DA SÉRIE.

Neste ponto do trabalho pretende-se conhecer melhor as características da série, para que seja possível uma boa escolha dos modelos a aplicar na próxima secção.

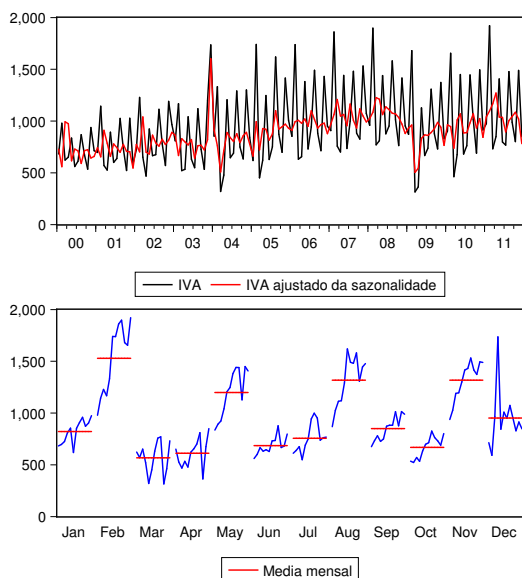


Figura 1: Cronograma e gráfico sazonal das receitas do IVA.

Na figura 1, o primeiro gráfico (cronograma da série), que descreve o comportamento desta (linha a preto) no período de 2000 a 2011¹. Da análise do gráfico vê-se um padrão de movimentos que se repetem anualmente, ou seja, o gráfico evidencia a presença de movimentos sazonais. Ainda no mesmo gráfico está presente a série dessazonalizada, isto é, sem o efeito da sazonalidade. A série dessazonalizada (linha a vermelho), mostra uma variação menor do valores observados. Quer pela série original quer pela série dessazonalizada é perceptível a existência de uma tendência, ainda que

pouco acentuada.

Ainda na figura 1, no segundo gráfico está decomposto do efeito sazonal, ou seja, o segundo gráfico demonstra o padrão sazonal ao longo do ano. Estes movimentos sazonais apresenta picos nos meses de fevereiro, maio, agosto e novembro, e baixas em março, abril e outubro. Estes picos e baixas são resultado da atividade das empresas e do prazo de pagamento do IVA por parte das empresas. O pico de fevereiro justifica-se pelo consumo elevado, habitual no mês de dezembro (época natalícia), cuja entrega do IVA ocorre 2 meses depois (fevereiro) e também pelo facto de o mês de fevereiro ser o mês de entrega do IVA do 4º trimestre do ano transato. Outro dois picos que, aparentemente têm um fator comum que justifique os picos são os picos de agosto e novembro. Ambos são o segundo mês após o termino do trimestre (2º e 3º trimestres), como para o apuramento mensal são o segundo mês dos meses de férias (junho e agosto), onde o consumo e a exportação (turismo) aumentam mais face aos outros meses.

Através da aplicação TRAMO², desenvolvida por [Gómez e Maravall (1994)], foi detetado a existência de dois *outliers*. Um na observação respeitante ao mes de dezembro de 2003. Neste caso estamos perante um *additive outlier*, assim sendo a série pode ser representada da seguinte forma,

$$Y_t = X_t + \omega_1 I_t^{(2003M12)} \quad (1)$$

onde X_t representa a série sem o efeito do *outlier* e

$$I_t^{(2003M12)} = \begin{cases} 0, & t \neq 2003M12 \\ 1, & t = 2003M12 \end{cases} \quad (2)$$

O outro *outlier* ocorre a observação respeitante ao mês de fevereiro de 2009. Neste caso estamos perante um *level shift outlier*, assim sendo a série pode ser representada da seguinte forma,

$$Y_t = Z_t + \omega_1 I_t^{(2003M12)} + \omega_2 I_t^{(2009M02)} \quad (3)$$

onde Z_t representa a série sem o efeito dos *outliers* e

$$I_t^{(2009M02)} = \begin{cases} 0, & t < 2009M02 \\ 1, & t \geq 2009M02 \end{cases} \quad (4)$$

Isto quer dizer que quando se for aplicar os modelos para previsão terá de se ter em conta esta informação.

¹Período definido para a modelização. O resto da amostra, como referido na introdução, será usado para avaliar a qualidade das previsões.

²Disponível no programa Eviews.

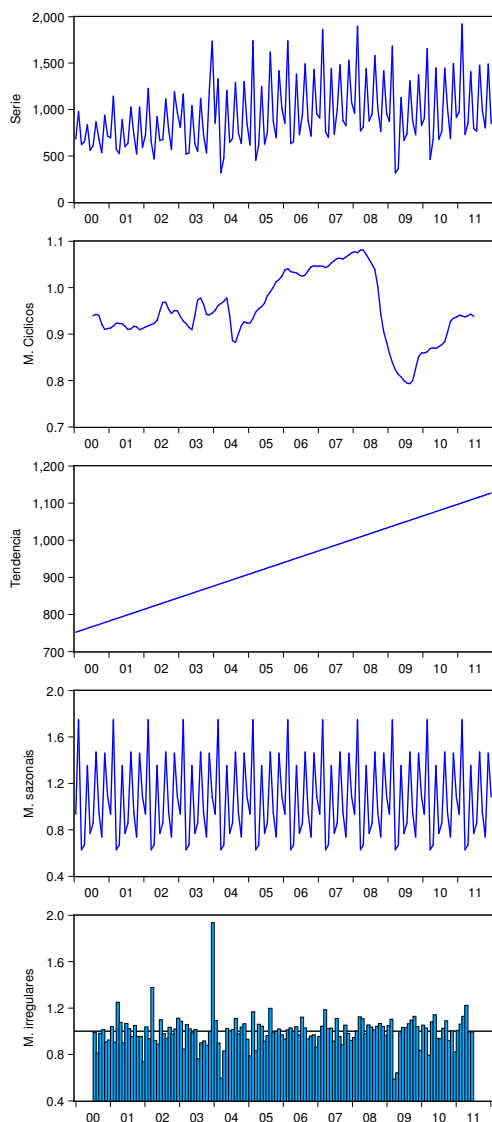


Figura 2: Decomposição da série através do método multiplicativo.

Com o objetivo de elaborar uma análise mais profunda da série, aplicou-se a técnica de decomposição clássica. Visto haver um aumento da sazonalidade com o aumento, ainda que ligeiro, da tendência, a técnica aplicada é aquela em que as componentes relacionam-se de forma multiplicativa. Os resultados desta decomposição estão evidenciados na figura 2. Os principais aspetos a realçar desta decomposição são: primeiro a tendência da série é crescente; segundo através da regressão da série sob o termo tendência, verifica-se que a série apresenta em média um crescimento mensal de 2,62 milhões de euros; terceiro parece existir ao longo do tempo um aumento da amplitude dos movimentos cíclicos; quarto a componente sazonal apresenta uma amplitude considerável, o

³Especificações do teste: Considerou-se um termo tendência e para a seleção dos desfasamentos aplicou-se o critério de Schwarz com 13 desfasamentos máximos.

que reforça a ideia de que a série apresenta uma sazonalidade forte e por fim os movimentos irregulares estão muito próximos de um à exceção, bastante clara, das observações que correspondem os outliers.

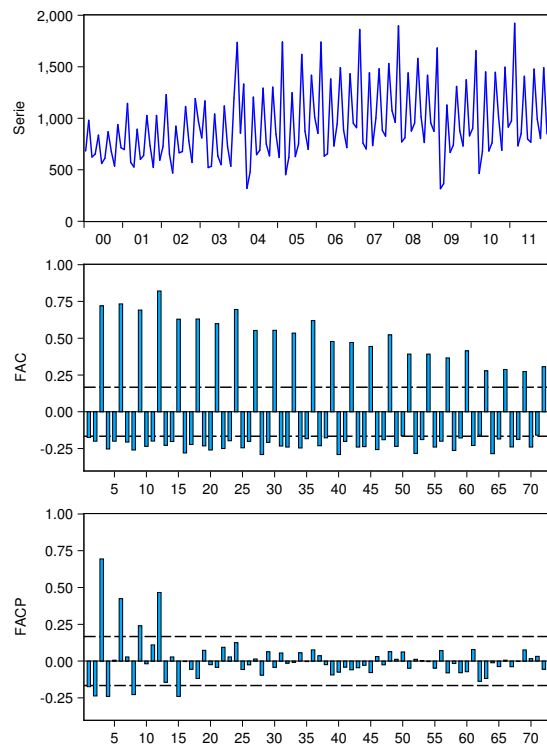


Figura 3: Cronograma, FAC e FACP da variável inicial.

Como se pode ver pelo primeiro gráfico da figura 3, no período inicial (ate 2003) a série apresenta uma variância menor de que a verificada após essa data. Este facto indica que a série não é estacionária em variância. No mesmo gráfico verifica-se, e como referido anteriormente, que a série apresenta uma tendência, o que implica que esta não seja estacionária em média. Isto quer dizer que a série não cumpre os requisitos para ser estacionária de segunda ordem ou em covariância. Para corroborar esta análise efetuou-se um teste de raiz unitária de *Dickey-Fuller*³, com a dimensão do teste (α) a 5%, onde se obteve o valor-p de 0.7159, o que é favorável à hipótese nula (não estacionaridade). Dadas as conclusões retiradas no paragrafo anterior, foi necessário transformar a variável. As transformações efetuadas foram: a aplicação do logaritmo natural e de seguida uma diferenciação sazonal ($\Delta Y_t = Y_t - Y_{t-12}$). As razões para aplicação da diferenciação sazonal em vez da diferenciação simples devem-se aos seguintes factos:

primeiro pela FAC verifica-se que tanto os primeiros *lags* (que indicam a estrutura não sazonal da série) como os *lags* sazonais demoram muito a atingir o valor de zero; segundo a série logaritimizada e com uma diferenciação simples, continuava a apresentar um decaimento muito lento para zero dos *lags* sazonais. Dados estes factos seria necessário sobre a diferenciação simples, aplicar uma diferenciação sazonal, para estacionarizar a série. Como se pode verá adiante, a aplicação somente da diferenciação sazonal à série logaritimizada, também transforma a série numa série estacionaria. Assim sendo para não sobre transformar a série decidi aplicar o logaritmo e a diferenciação sazonal. Obteve-se a série caracterizada, pelo seu cronograma, função de autocorrelação (FAC) e função autocorrelação parcial (FACP), na figura 4.

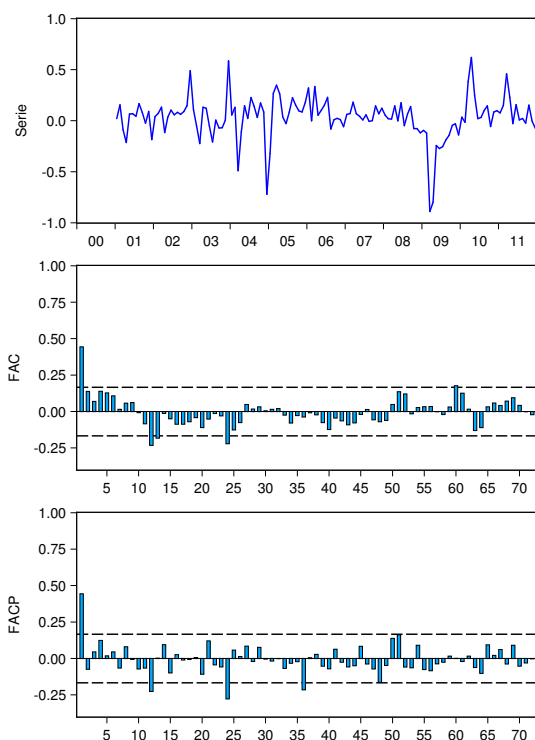


Figura 4: Cronograma, FAC e FACP da variável transformada.

Quer pelo que é evidenciado no primeiro gráfico da figura 4, quer pelo valor-p (0,0000) do teste de raiz unitária efetuado à série transformada, há evidencias para se rejeitar o cenário de não estacionaridade, ou seja, não há indícios de que a série transformada não seja estacionaria em covariância.

III. MODELIZAÇÃO.

Neste ponto irá ser descrita a metodologia usada para efetuar as previsões, que serão analisadas no próximo ponto, e a forma de analisar a qualidade da previsão. Em termos gerais, o indicador de avaliação da previsão a ser utilizado será o Erro Absoluto Médio (EAM) e a metodologia aplicada divide-se em: alisamento exponencial, modelo ARIMA e combinação de previsões.

I. Avaliação dos erros de previsão.

Seja Y_t uma série temporal, onde t representa o período em que ocorre a observação e \hat{Y}_{t+m} a previsão da série temporal (Y_t) para o período $t + m$ a m passos (i.e. a m períodos de distancia da realização da série). A medida escolhida para a análise da qualidade de previsão, isto é, a estatística escolhida para medir o erro de previsão é o Erro Absoluto Médio (EAM). A estatística é obtida da seguinte forma

$$\text{EAM} = \frac{1}{n} \sum_{i=1}^n |\hat{Y}_i - Y_i| \quad (5)$$

onde n é o número de previsões efetuadas. A principal justificação para a utilização do EAM em vez do Erro Quadrático Médio (medida mais usual) deve-se ao facto da interpretação do valor assumido pelo EAM ser mais simples, visto a unidade do EAM ser a mesma da variável. E não o quadrado desta, como acontece no EQM.

II. Alisamento Exponencial.

Existe uma gama variada de métodos de alisamento exponencial, cada um indicado para series temporais distintas [ver [Gardner (1985)] ou a versão mais recente [Gardner (2006)]].

Dadas as características da série como a existência de tendência e sazonalidade (com os movimento sazonais a aumentar ao mesmo tempo que a tendência) o modelo de alisamento exponencial escolhido é o modelo de Holt-Winters com tendência aditiva amortecida e sazonalidade multiplicativa. Este modelo é composto do três equações de atualização uma para o nível ($a(t)$), outra para a tendência ($b(t)$) da série e outra para a sazonalidade ($S(t)$) e

a equação de previsão.

$$a(t) = \alpha \frac{Y_t}{S(t-s)} + (1-\alpha)(a(t-1) + \phi b(t-1)) \quad (6)$$

$$b(t) = \beta(a(t) - a(t-1)) + (1-\beta)\phi b(t-1) \quad (7)$$

$$S(t) = \gamma \frac{Y_t}{a(t)} + (1-\gamma)S(t-s) \quad (8)$$

$$\hat{Y}_{t+m} = (a(t) + b(t) \sum_{i=1}^m \phi^i) S(t+m-s). \quad (9)$$

com α, β, γ e $\phi \in [0, 1]$. Se o $\phi = 1$ temos o modelo de Holt-Winter multiplicativo. Para a aplicação deste modelo é necessário efetuar a iniciação das equações, isto é, definir o valor para $a(1), b(1)$ e $\{S(1), \dots, S(s)\}$. Os valores iniciais são dados por:

$$a(s) = \frac{1}{s} \sum_{i=1}^s Y_i \quad (10)$$

$$b(s) = \frac{1}{s^2} \left(\sum_{i=s+1}^{2s} Y_i - \sum_{i=1}^s Y_i \right) \quad (11)$$

$$S(1) = \frac{Y_1}{a(s)}, \dots, S(s) = \frac{Y_s}{a(s)} \quad (12)$$

Este modelo vai ser aplicado desde a 1 passo até 12 passos, isto quer dizer que se vai construir um modelo para prever o valor da série no período seguinte, outro modelo para prever o valor da série dois momentos à frente da altura em que é feita a previsão, etc até ao último modelo onde se vai prever o valor da série a 12 momentos à frente. Para a escolha dos parâmetros das equações de atualização, para cada um dos 12 modelos, α, β, γ e ϕ vou usar a escolha ótima através da aplicação Solver do programa Excel. Esta escolha ótima vai ter como objetivo, a escolha da combinação de parâmetros que minimiza o EAM, para parte do período de modelização (de janeiro de 2002 a dezembro de 2011).

III. Modelo ARIMA.

Um outro modelo usado para a previsão da série é o modelo ARIMA (ver [Box et al. (1994)]). O modelo ARIMA na sua construção geral é constituído por um polinómio autoregressivo ($\phi(B)$) e um de médias móveis ($\theta(B)$) mais os polinómios autoregressivos e de médias móveis sazonais ($\Phi(B)$ e $\Theta(B)$), ou seja,

$$\Phi(B)\phi(B)(1-B)^d(1-B^D)(Y_t - \mu) = \Theta(B)\theta(B)\zeta_t \quad (13)$$

onde B é o operador atraso, $\Phi(B) = (1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_P B^P)$, $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$, μ representa a média da série, $\Theta(B) = (1 - \Theta_1 B - \Theta_2 B^2 - \dots - \Theta_Q B^Q)$ e $\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$. Uma forma a simples de representar este modelo é SARIMA(p, d, q)(P, D, Q) $_s$, onde p é a ordem do polinómio autoregressivo, d é o número de diferenciações simples aplicadas à série, q é a ordem do polinómio de médias móveis, P é a ordem do polinómio autoregressivo sazonal, D é o número de diferenciações sazonais aplicadas à série e Q é a ordem do polinómio de médias móveis sazonal. A definição dos valores inteiros de p, q, d, P, Q e D é feita através da análise das FAC e FACP da série. Após esta definição o modelo será estimado e avaliado.

A avaliação da qualidade do modelo é feita a dois tempos. Primeiro são efetuados testes ao modelo com o objetivo de avaliar a qualidade das estimativas. Estes testes são: testes de significância ao parâmetros do modelo, o teste de significância a cada lag da FAC e FACP dos resíduos do modelo e o teste Ljung-Box [Ljung e Box (1978)] (testa da nulidade conjunta dos lags da FAC e FACP dos resíduos do modelo). Em segundo lugar, caso passem nos testes anteriores serão comparados ao nível do critério de seleção Schwartz (ver [Schwarz (1978)]).

IV. Combinação de previsões.

Quando não se consegue atingir o modelo ideal para a previsão de uma série, pode ser útil a combinação de previsões de diferentes modelos com vista a uma melhor previsão. Segundo [Makridakis e Winkler (1983)] a combinação de previsões pode produzir menos erros que a utilização de um modelo isoladamente. Nesse mesmo artigo, os autores para combinar as previsões utilizam uma média simples destas, ou seja,

$$\hat{Y}_{t+m}^c = \frac{1}{n} \sum_{i=1}^n \hat{Y}_{t+m}^i \quad (14)$$

onde \hat{Y}_{t+m}^c representa a previsão resultante da combinação dos modelos para o instante $t+m$ e \hat{Y}_{t+m}^i é a previsão do modelo i para o instante $t+m$. No mesmo ano, os mesmos autores (ver [Winkler e Makridakis (1983)]) testam a eficácia de fazer uma média ponderada das previsões dos modelos em vez de uma média simples. Um dos casos testados usaram o Erro Quadrático Médio (EQM) para ponderar inversamente as previsões, ou seja, somaram o EQM de todos só modelos e na média

atribuíram a cada um dos modelos o peso inversamente proporcional à contribuição para o total do EQM. Neste trabalho procurar-se-á efetuar a mesma experiência só que usando o EAM. Assim sendo temos,

$$\hat{Y}_{t+m}^c = \frac{1}{(k-1)M} \sum_{i=1}^k (M - EAM^i) \hat{Y}_{t+m}^i \quad (15)$$

onde $M = \sum_{i=1}^k EAM^i$, ou seja, M é a soma do EAM de todos os modelos e EAM^i é o Erro Absoluto Médio do modelo i .

IV. RESULTADOS.

Esta secção está dividida em três partes: a primeira onde se irá apresentar os modelos estimados, a segunda onde se apresenta os resultados da previsão dos modelos para o primeiro período de teste e uma terceira parte onde se aplicará a combinação de previsões e onde se analisa os resultados das previsões efetuadas para o segundo período de teste.

I. Resultados da modelização.

Através da otimização, os resultados da aplicação do modelo de Holt-Winters com tendência aditiva amortecida e sazonalidade multiplicativa estão descritos na tabela 1.

Previsão a	Coeficientes			
	α	β	γ	ϕ
1 passo	0,134	0,000	0,677	0,943
2 passos	0,109	0,000	0,513	0,938
3 passos	0,124	0,000	0,510	0,939
4 passos	0,063	0,000	0,571	0,913
5 passos	0,070	0,000	0,564	0,899
6 passos	0,085	0,000	0,553	0,897
7 passos	0,068	0,000	0,614	0,888
8 passos	0,068	0,000	0,662	0,880
9 passos	0,087	0,000	0,704	0,880
10 passos	0,085	0,000	0,744	0,875
11 passos	0,094	0,000	0,660	0,862
12 passos	0,000	0,092	0,764	0,877

Tabela 1: Coeficientes ótimos do modelo de Holt-Winters com tendência aditiva amortecida e sazonalidade multiplicativa.

Os doze modelos apresentam coeficientes muito semelhantes⁴, com particular interesse num

β nulo em quase todos os todos os modelos. A consequência do β ser nulo e de o ϕ inferior a 1, é a da componente tendência tender para 0. Basicamente, à medida que o momento, em que se efetua a previsão, se distancia da inicialização das equações de atualização, a equação de atualização da tendência aproxima-se de zero. Logo as previsões efetuadas baseiam-se sobretudo na componente do nível e da sazonalidade da série.

Após a análise da FAC e FACP da série transformada, decidi estimar os seguintes modelos, tendo já em conta os *outliers*,

$$(1 - \Phi_{12}B^{12})X_t = (1 - \Theta_{24}B)(1 - \theta_1B)\zeta_t + \omega_1 I_t^{(2003M12)} + \omega_2 I_t^{(2009M02)} \quad (16)$$

$$(1 - \Phi_{24}B^{24})(1 - \phi_1B)X_t = (1 - \Theta_{12}B^{12})\zeta_t + \omega_1 I_t^{(2003M12)} + \omega_2 I_t^{(2009M02)} \quad (17)$$

onde $X_t = (1 - B^{12})(Y_t - c)$.

Como se pode verificar pela tabela 2 ambos os modelos passam nos testes efetuados, incluindo a significância individual dos *lags* da FAC e FACP (ver Anexo). É de realçar que ambos os modelos apresentam valores-p elevados nos testes estatísticos, o que quer dizer que as estatísticas teste observadas são bastante favoráveis às hipóteses nulas dos testes. A exceção ocorre no teste de normalidade dos resíduos, onde ambos os modelos não passam no teste.

II. Previsão para o primeiro período de teste.

Com base nas catorze modelizações da série (12 alisamentos exponenciais e 2 modelos ARIMA), foi efetuada a previsão da série. Para cada uma das modelizações foi calculados o EAM, que se apresenta na tabela 3.

Os modelos de previsão baseados no alisamento exponencial apresentam melhores resultados que os modelos SARIMA. Um aspeto a salientar é a proximidade, em termos de qualidade, dos alisamento de Holt-Winters independentemente da distancia a que é feita a previsão.

Juntando os resultados da tabela 2 e da tabela 3 verifica-se que o modelo que se ajusta melhor estatisticamente aos dados, não é, neste caso, o melhor modelo para prever a série. Isto verifica-se, pelo facto de o modelo SARIMA 1 (SARIMA(0,0,1)(1,1,2)₁₂) apresen-

⁴Os coeficientes dos modelos serem iguais, não leva a que as previsões obtidas de cada modelo sejam iguais. Isto acontece porque o momento em que são efetuadas a previsões para aquele ano são diferentes.

Modelo 1- SARIMA(0,0,1)(1,1,2) ₁₂ com $\Theta_{12} = 0$			Resultados gerais do modelo	
Parâmetro	Coefficiente	Valor-p		
c	0.0602	0.0000	Teste Ljung-Box aos primeiros 48 lags (valor-p)	0.9057
θ	-0.2903	0.0019	Teste Normalidade dos resíduos de Jarque-Bera (valor-p)	0.0000
Φ	-0.5558	0.0000	Critério de informação de Schwarz	-1.2071
Θ_{24}	0.8885	0.0000		
ω_1	0.8474	0.0000		
ω_2	-0.2430	0.0000		

Modelo 2- SARIMA(1,0,0)(2,1,1) ₁₂ com $\Phi_{12} = 0$			Resultados gerais do modelo	
Parâmetro	Coefficiente	Valor-p		
c	0.0683	0.0000	Teste Ljung-Box aos primeiros 16 lags (valor-p)	0.8983
ϕ	0.1638	0.0994	Teste Normalidade dos resíduos de Jarque-Bera (valor-p)	0.0000
Φ_{24}	-0.3661	0.0001	Critério de informação de Schwarz	-1.1873
Θ	0.9316	0.0000		
ω_1	0.7943	0.0000		
ω_2	-0.2876	0.0000		

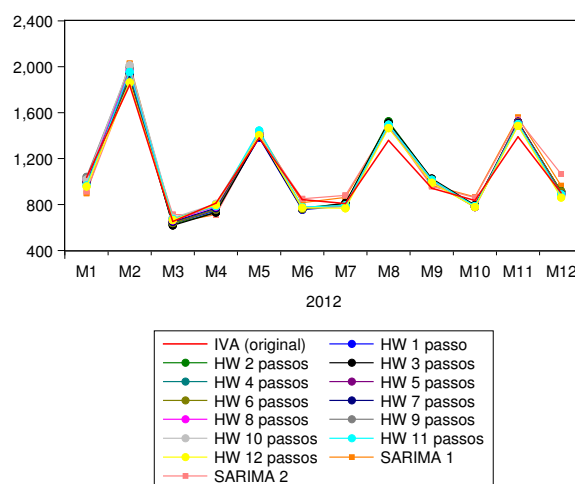
Tabela 2: Resultados da aplicação do modelo ARIMA.

tar melhores resultados nos testes estatísticos e no critério de informação do que o modelo SARIMA 2 (SARIMA(1,0,0)(2,1,1)₁₂), mas o modelo SARIMA 2 apresenta um EAM mais baixo que o SARIMA 1.

Modelo	Erro Absoluto Médio
HW 1 passo	61,177
HW 2 passos	60,260
HW 3 passos	62,065
HW 4 passos	58,378
HW 5 passos	57,139
HW 6 passos	58,110
HW 7 passos	56,567
HW 8 passos	56,670
HW 9 passos	59,148
HW 10 passos	63,666
HW 11 passos	62,838
HW 12 passos	51,125
SARIMA 1	82,903
SARIMA 2	76,046

Tabela 3: Resultados da previsão para o primeiro período de teste.

Da análise ao gráfico 5 pode-se afirmar que o padrão sazonal foi bem capturado pelos modelos. Geralmente como se pode ver pelo gráfico 5 a grande maioria dos modelos aproxima-se do valor observado.


Figura 5: Previsão individual para o primeiro período de teste.

III. Previsão para o segundo período de teste.

Foi ainda, para o ano de 2013, efetuada a combinação das previsões, com vista a testar se esta traria ou não melhores resultados que a previsão individual para o mesmo período. Como exposto anteriormente (ver [Combinação de previsões](#)), as formas de combinação das previsões são dadas por:

• combinação 1

$$\hat{Y}_{t+m}^c = \frac{1}{14} (\hat{Y}_{t+m}^{HW1} + \hat{Y}_{t+m}^{HW2} + \hat{Y}_{t+m}^{HW3} + \hat{Y}_{t+m}^{HW4} + \hat{Y}_{t+m}^{HW5} + \hat{Y}_{t+m}^{HW6} + \hat{Y}_{t+m}^{HW7} + \hat{Y}_{t+m}^{HW8} + \hat{Y}_{t+m}^{HW9} + \hat{Y}_{t+m}^{HW10} + \hat{Y}_{t+m}^{HW11} + \hat{Y}_{t+m}^{HW12} + \hat{Y}_{t+m}^{SARIMA1} + \hat{Y}_{t+m}^{SARIMA2}) \quad (18)$$

• combinação 2

$$\hat{Y}_{t+m}^c = \frac{1}{13M} \left[(M - EAM^{H1}) \hat{Y}_{t+m}^{H1} + (M - EAM^{H2}) \hat{Y}_{t+m}^{H2} + (M - EAM^{H3}) \hat{Y}_{t+m}^{H3} + (M - EAM^{H4}) \hat{Y}_{t+m}^{H4} + (M - EAM^{H5}) \hat{Y}_{t+m}^{H5} + (M - EAM^{H6}) \hat{Y}_{t+m}^{H6} + (M - EAM^{H7}) \hat{Y}_{t+m}^{H7} + (M - EAM^{H8}) \hat{Y}_{t+m}^{H8} + (M - EAM^{H9}) \hat{Y}_{t+m}^{H9} + (M - EAM^{H10}) \hat{Y}_{t+m}^{H10} + (M - EAM^{H11}) \hat{Y}_{t+m}^{H11} + (M - EAM^{H12}) \hat{Y}_{t+m}^{H12} + (M - EAM^{SARIMA1}) \hat{Y}_{t+m}^{SARIMA1} + (M - EAM^{SARIMA2}) \hat{Y}_{t+m}^{SARIMA2} \right] \quad (19)$$

onde $M = \sum_{i=1}^k EAM^i$, ou seja, M é a soma do EAM de todos os modelos.

Modelo	Ponderadores Baseado no EAM
H1 passo	7,149%
H2 passos	7,157%
H3 passos	7,141%
H4 passos	7,174%
H5 passos	7,185%
H6 passos	7,176%
H7 passos	7,190%
H8 passos	7,189%
H9 passos	7,167%
H10 passos	7,127%
H11 passos	7,134%
H12 passos	7,238%
SARIMA 1	6,956%
SARIMA 2	7,017%

Tabela 4: Ponderadores para a combinação 2.

Na tabela 4 estão evidenciados os ponderadores para as combinação 2. Os ponderadores são muito

semelhantes entre si, contudo há ligeiras diferenças. As diferenças mais notórias residem na menor importância dada ao modelo SARIMA 1. Apesar das ligeiras diferenças, há uma proximidade dos valores, o que indica que os resultados desta combinação serão bastante semelhantes aos da combinação 1.

Modelo	Erro Absoluto Médio
HW 1 passo	79,417
HW 2 passos	87,160
HW 3 passos	87,442
HW 4 passos	85,790
HW 5 passos	83,512
HW 6 passos	83,930
HW 7 passos	85,565
HW 8 passos	82,505
HW 9 passos	82,116
HW 10 passos	82,529
HW 11 passos	82,618
HW 12 passos	78,696
SARIMA 1	88,471
SARIMA 2	115,143
Combinação 1	79,833
Combinação 2	79,913

Tabela 5: Resultados da previsão para o segundo período de teste.

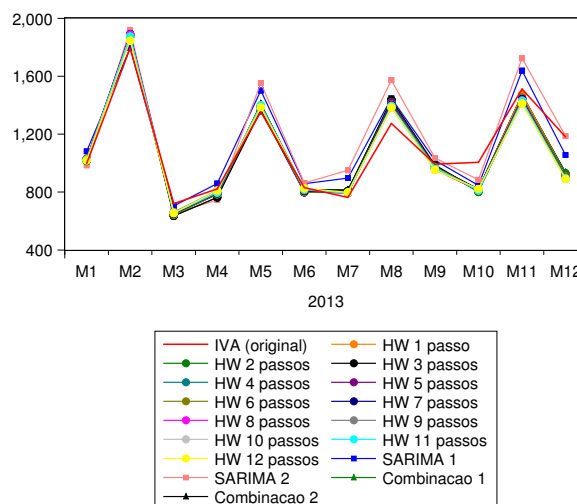


Figura 6: Previsão individual e combinada para o segundo período.

A tabela 5 apresenta os EAM da previsão para o segundo período, ou seja, para o ano de 2013, onde se observa a boa capacidade de previsão dos modelos de HW a 1 passo e a 12 passo. Em termos de combinação de previsões, a que apresenta os

melhores resultados é a combinação 1 com resultados muito semelhantes à combinação 2. Quando se compara as previsões obtidas através das combinações e dos modelos, verifica-se que o top-4 é preenchido pelos modelos dois modelos de HW referidos anteriormente, fechando o top com as combinações.

Como se pode ver pela figura 6, não parece existir um modelo coerente para a previsão da série daí que entre os melhores preditores estejam as combinações.

V. CONCLUSÃO.

Com a elaboração deste estudo sobre a modelação e previsão das receitas do IVA, foi possível retirar algumas conclusões.

Em primeiro lugar, é de realçar a qualidade de modelação e previsão da série através da utilização de modelos de alisamento exponencial. Estes modelos, de aplicação mais simples que os modelos ARIMA, obtiveram em qualquer um dos períodos da amostra usados para testes, resultados superiores aos modelos ARIMA.

Em segundo lugar, os resultados da combinação de previsões seguiram a mesma linha dos melhores modelos de previsão. Isto quer dizer, a combinação de previsões não conseguiu obter melhores resultados, que os melhores modelos. Este facto acontece pela pouca diversidade de previsões. Por outras palavras, apesar dos modelos utilizados a diferença entre as previsões, não é tão significativa que permita a existência de compensações entre previsões por excesso e por defeito.

REFERÊNCIAS

- [Box et al. (1994)] Box, G., Jenkins, G., and Reinsel, G. 1994. *Time series analysis: Forecasting and control*, 3rd Ed., Prentice-Hall, Englewood Cliffs, N.J.
- [Gardner (1985)] Gardner Jr., E 1985, 'Exponential Smoothing: The State of the Art', *Journal Of Forecasting*, 4, 1, pp. 1-28.
- [Gardner (2006)] Gardner Jr., E 2006, 'Exponential smoothing: The state of the art—Part II', *International Journal Of Forecasting*, 22, pp. 637-666.
- [Gómez e Maravall (1994)] Gómez, V, & Maravall, A 1994, 'Estimation, prediction, and interpolation for nonstationary series with the Kalman filter', *Journal Of The American Statistical Association*, 89, 426, p. 611.
- [Ljung e Box (1978)] Ljung, G, & Box, G 1978, 'On a Measure of Lack of Fit in Time Series Models', *Biometrika*, 2, p. 297.
- [Makridakis e Winkler (1983)] Makridakis, S, & Winkler, R 1983, 'AVERAGES OF FORECASTS: SOME EMPIRICAL RESULTS', *Management Science*, 29, 9, pp. 987-996.
- [Schwarz (1978)] Schwarz, G 1978, 'Estimating the Dimension of a Model', *The Annals of Statistics*, 2, p. 461.
- [Winkler e Makridakis (1983)] Winkler, R, & Makridakis, S 1983, 'The Combination of Forecasts', *Journal of the Royal Statistical Society. Series A (General)*, 2, p. 150.

VI. ANEXO.

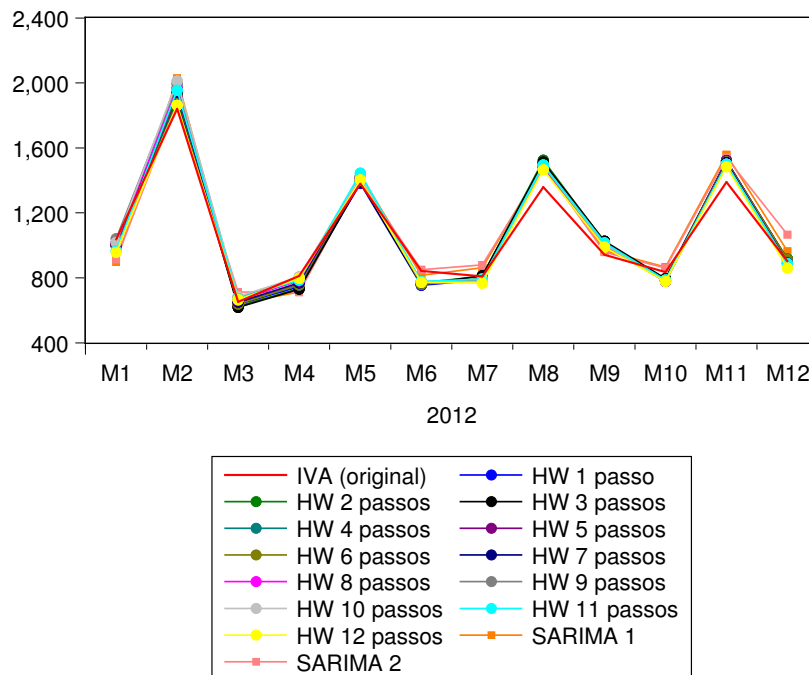


Figura 7: Previsão individual para o ano de 2012.

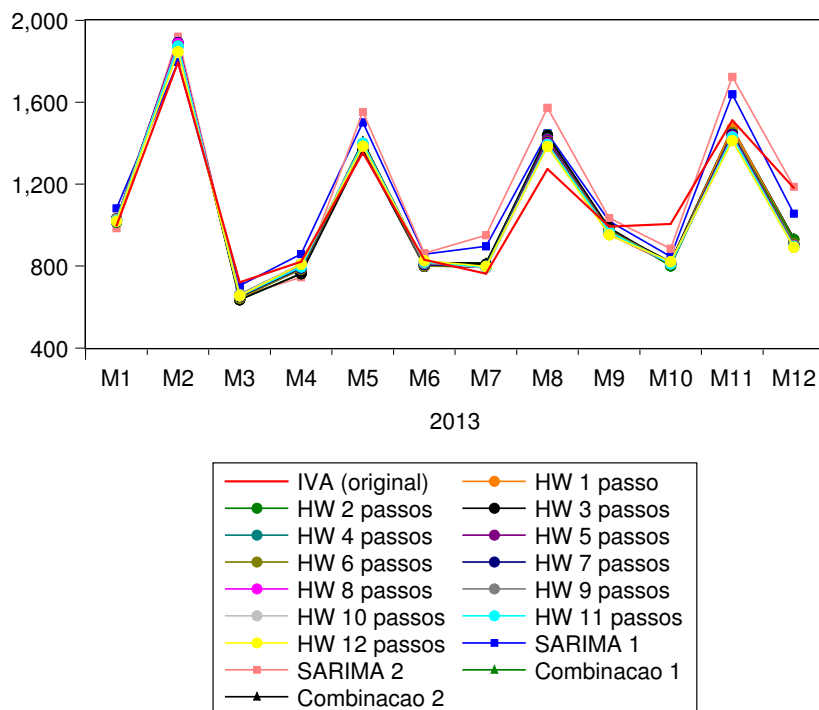


Figura 8: Previsão individual e combinada para o ano de 2013.

Dependent Variable: D(LOG(IVA),0,12)
 Method: Least Squares
 Date: 01/05/15 Time: 01:05
 Sample (adjusted): 2002M01 2011M12
 Included observations: 120 after adjustments
 Convergence achieved after 12 iterations
 MA Backcast: 1999M12 2001M12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.060237	0.009249	6.512745	0.0000
D(AO03M12,0,12)	0.847404	0.159214	5.322420	0.0000
D(LS09M02,0,12)	-0.242955	0.050188	-4.840879	0.0000
AR(12)	-0.555820	0.075859	-7.326993	0.0000
MA(1)	0.290294	0.091124	3.185703	0.0019
SMA(24)	-0.888460	0.030031	-29.58450	0.0000
R-squared	0.690062	Mean dependent var		0.036087
Adjusted R-squared	0.676468	S.D. dependent var		0.211757
S.E. of regression	0.120447	Akaike info criterion		-1.346503
Sum squared resid	1.653860	Schwarz criterion		-1.207128
Log likelihood	86.79015	Hannan-Quinn criter.		-1.289902
F-statistic	50.76304	Durbin-Watson stat		2.003367
Prob(F-statistic)	0.000000			
Inverted AR Roots	.92-.25i	.92+.25i	.67-.67i	.67+.67i
	.25-.92i	.25+.92i	-.25+.92i	-.25-.92i
	-.67-.67i	-.67+.67i	-.92-.25i	-.92+.25i
Inverted MA Roots	1.00	.96-.26i	.96+.26i	.86+.50i
	.86-.50i	.70+.70i	.70-.70i	.50-.86i
	.50+.86i	.26-.96i	.26+.96i	.00-1.00i
	-.00+1.00i	-.26-.96i	-.26+.96i	-.29
	-.50-.86i	-.50+.86i	-.70+.70i	-.70-.70i
	-.86-.50i	-.86+.50i	-.96-.26i	-.96+.26i
	-1.00			

Figura 9: Modelo SARIMA 1.

Dependent Variable: D(LOG(IVA),0,12)
 Method: Least Squares
 Date: 01/05/15 Time: 01:09
 Sample (adjusted): 2003M02 2011M12
 Included observations: 107 after adjustments
 Convergence achieved after 24 iterations
 MA Backcast: 2002M02 2003M01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.068302	0.007014	9.737474	0.0000
D(AO03M12,0,12)	0.794344	0.122640	6.477024	0.0000
D(LS09M02,0,12)	-0.287636	0.041365	-6.953646	0.0000
AR(1)	0.163810	0.098503	1.663000	0.0994
SAR(24)	-0.366096	0.088528	-4.135365	0.0001
MA(12)	-0.931577	0.019220	-48.46926	0.0000
R-squared	0.709473	Mean dependent var		0.028304
Adjusted R-squared	0.695091	S.D. dependent var		0.218523
S.E. of regression	0.120665	Akaike info criterion		-1.337151
Sum squared resid	1.470572	Schwarz criterion		-1.187272
Log likelihood	77.53755	Hannan-Quinn criter.		-1.276392
F-statistic	49.32885	Durbin-Watson stat		1.904110
Prob(F-statistic)	0.000000			
Inverted AR Roots	.95-.13i	.95+.13i	.89+.37i	.89-.37i
	.76+.58i	.76-.58i	.58+.76i	.58-.76i
	.37-.89i	.37+.89i	.16	.13-.95i
	.13+.95i	-.13-.95i	-.13+.95i	-.37-.89i
	-.37+.89i	-.58-.76i	-.58+.76i	-.76-.58i
	-.76+.58i	-.89-.37i	-.89+.37i	-.95-.13i
	-.95+.13i			
Inverted MA Roots	.99	.86+.50i	.86-.50i	.50-.86i
	.50+.86i	.00+.99i	-.00-.99i	-.50+.86i
	-.50-.86i	-.86+.50i	-.86-.50i	-.99

Figura 10: SARIMA 2.

Date: 01/09/15 Time: 03:07
 Sample: 2000M01 2011M12
 Included observations: 120
 Q-statistic probabilities adjusted for 3 ARMA terms and 2 dynamic regressors

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob*	
		1	-0.016	-0.016	0.0308	
		2	-0.007	-0.007	0.0365	
		3	0.139	0.139	2.4514	
		4	0.040	0.045	2.6556	0.103
		5	0.115	0.120	4.3246	0.115
		6	0.009	-0.004	4.3353	0.227
		7	-0.032	-0.043	4.4701	0.346
		8	-0.011	-0.049	4.4851	0.482
		9	-0.069	-0.086	5.1141	0.529
		10	0.045	0.038	5.3887	0.613
		11	-0.072	-0.063	6.0799	0.638
		12	-0.019	0.014	6.1305	0.727
		13	-0.122	-0.128	8.1782	0.611
		14	0.013	0.042	8.2004	0.695
		15	-0.025	-0.036	8.2902	0.762
		16	-0.136	-0.096	10.903	0.619
		17	-0.060	-0.064	11.406	0.654
		18	-0.054	-0.041	11.824	0.692
		19	0.102	0.144	13.324	0.649
		20	-0.126	-0.125	15.660	0.548
		21	-0.049	0.006	16.013	0.592
		22	-0.037	-0.099	16.222	0.642
		23	0.063	0.121	16.824	0.664
		24	-0.103	-0.175	18.447	0.621
		25	-0.103	-0.077	20.069	0.579
		26	-0.014	-0.054	20.100	0.636
		27	-0.008	0.020	20.110	0.690
		28	-0.028	-0.007	20.236	0.734
		29	0.016	-0.037	20.276	0.778
		30	-0.044	0.019	20.595	0.805
		31	0.015	-0.038	20.634	0.840
		32	0.034	0.083	20.822	0.866

*Probabilities may not be valid for this equation specification.

Figura 11: Output do Eviews do teste de Ljung e Box aplicados ao modelo SARIMA 1.

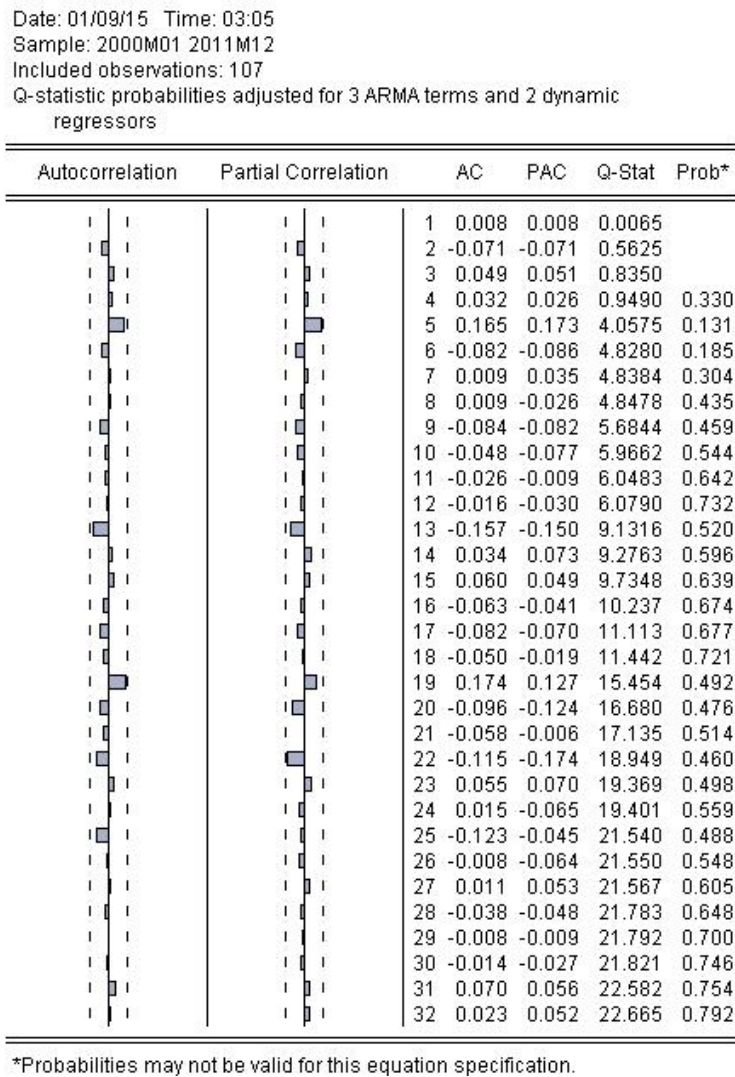


Figura 12: Output do Eviews do teste de Ljung e Box aplicados ao modelo SARIMA 2.

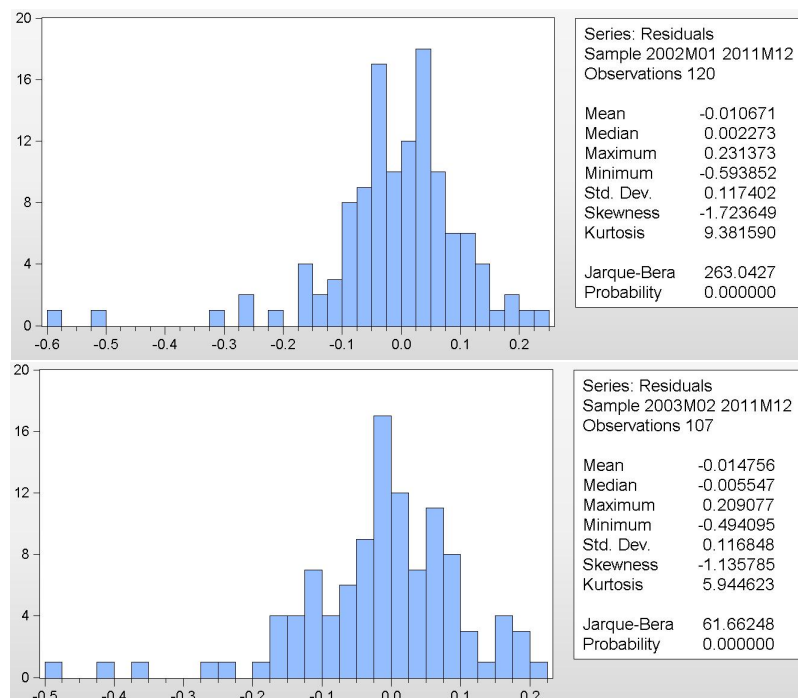


Figura 13: Output do Eviews do teste de normalidade aos resíduos aplicados ao modelo SARIMA 1 e SARIMA 2.