

# UMA APLICAÇÃO DE MODELOS DE SÉRIES TEMPORAIS NA PREVISÃO DA DEMANDA POR TRANSPORTE AÉREO DE PASSAGEIROS

**Marcial Alexandre Marazzo da Silva**

Departamento de Aviação Civil  
Divisão de Planejamento

**Luiz Henrique da Conceição Leal**

Departamento de Aviação Civil  
Divisão de Planejamento

**José Francisco Moreira Pessanha**

Universidade do Estado do Rio de Janeiro  
Instituto de Matemática e Estatística

## RESUMO

O presente trabalho tem como objetivo apresentar a utilização de uma modelagem de séries temporais do tipo Box & Jenkins na previsão da demanda por transporte aéreo regular, para o segmento doméstico brasileiro. A metodologia ora apresentada é aplicada com o intuito de se obter uma visão prospectiva de curto prazo para o comportamento da variável passageiros-quilômetros transportados, utilizando os modelos ARIMA (Auto Regressivo – Integrado – Média Móvel).

## ABSTRACT

The proposal of the present paper is to show a times series modeling applied to scheduled domestic air transport flights demand in Brazil. The methodology is used to obtain a short term forecast for the passenger-kilometer variable, using an ARIMA model (Autoregressive Integrated Moving Average).

## 1. INTRODUÇÃO

Praticamente todas as empresas, sejam elas de pequeno, médio ou grande porte, estatais, nacionais privadas ou multinacionais, necessitam planejar seus recursos de produção, vendas ou serviços, levando em consideração condições futuras incertas. Dentro deste contexto a utilização de métodos científicos de previsão se torna fundamental como ferramenta de auxílio no processo de decisão, uma vez que para o transporte aéreo, devido ao alto custo de operação, a formulação de um planejamento incorreto incorre em prejuízos irrecuperáveis.

A modelagem de séries temporais possibilita descrever um processo estocástico, utilizando valores passados da variável de interesse, baseando-se na idéia de que observações passadas contêm informações sobre o padrão de comportamento da série temporal. Entre as diversas metodologias utilizadas para este fim, destacam-se os modelos de Box & Jenkins (1994) e os modelos de Amortecimento Exponencial (Morettin e Toloi, 1987). Segundo Goodrich (1989), quando a série analisada é pequena, i.e., menos de 40 observações, a melhor escolha consiste em usar algum modelo de amortecimento exponencial. Esta escolha é justificada pelo fato de que em séries pequenas a estimativa da função de autocorrelação é bastante instável.

O objetivo do presente trabalho é discutir e aplicar as técnicas de modelagem de séries temporais, utilizando os métodos de Box & Jenkins, para identificar o padrão de comportamento da demanda por transporte aéreo no mercado doméstico brasileiro e distinguir este padrão de qualquer ruído que possa estar contido nas observações, permitindo assim, realizar estimativas consistentes para valores futuros da série.

A série histórica analisada no presente trabalho consiste nos registros mensais da variável passageiros-quilômetros pagos transportados, no período de janeiro de 2000 a fevereiro de 2005, publicados mensalmente pelo Departamento de Aviação Civil (DAC) através de seu endereço eletrônico [www.dac.gov.br](http://www.dac.gov.br).

Este artigo está organizado em cinco seções. A seguir, na seção 2 tem-se uma breve introdução aos modelos de Box & Jenkins. Na seção 3, decreve-se o processo de identificação do modelo Box & Jenkins que melhor representa a série passageiros-quilômetros transportados no período de janeiro de 2000 a setembro de 2004. Na seção 4 são apresentadas as previsões para o período de outubro de 2004 a fevereiro de 2005, obtidas pelo melhor modelo identificado. Finalmente, na seção 5 são resumidas as principais conclusões do trabalho.

## 2. ASPECTOS METODOLÓGICOS

Tendo-se como base uma determinada série temporal, gerada por um processo estocástico estacionário (processo invariante com relação ao tempo), onde o valor atual é dado por  $y_t$ , o processo estocástico gerador da série pode ser descrito como segue:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (1)$$

ou, escrevendo de outra forma:

$$(1 - \phi_1 B - \dots - \phi_p B^p) y_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \quad (2)$$

onde  $B$  é o operador de defasagem ( $By_t = y_{t-1}$ ),  $\phi$  e  $\theta$  são os parâmetros do modelo e  $\varepsilon_t$  um termo de erro com distribuição normal, média nula, variância constante e não autocorrelacionado (ruído branco). As defasagens da variável  $y_t$  representam a parte autoregressiva do modelo (AR), enquanto as defasagens do erro representam a parte média móvel (MA). O modelo apresentado acima é denominado ARMA(p,q), onde p representa a ordem de defasagem do termo autoregressivo e q a ordem de defasagem do termo de média móvel. No caso dos modelos puramente sazonais tem-se que:

$$(1 - \Phi_1 B^S - \dots - \Phi_P B^{PS}) y_t = (1 - \Theta_1 B^S - \dots - \Theta_Q B^{QS}) \varepsilon_t \quad (3)$$

onde  $\Phi$  e  $\Theta$  são os parâmetros do modelo sazonal, e  $S$  é o período sazonal (por exemplo,  $S = 12$ , para séries mensais, e  $S = 4$ , para séries trimestrais). No caso geral, as séries temporais podem apresentar componentes sazonais e não sazonais. Neste caso, o modelo é representado da seguinte forma:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - \Phi_1 B^S - \dots - \Phi_P B^{PS})y_t = (1 - \theta_1 B - \dots - \theta_q B^q)(1 - \Theta_1 B^S - \dots - \Theta_Q B^{QS})\varepsilon_t \quad (4)$$

Assim, tem-se um modelo misto ARMA(p,q)(P,Q)s, onde  $p$  e  $q$  referem-se, respectivamente, às ordens autoregressiva e de média móvel da componente não sazonal; enquanto que  $P$  e  $Q$  referem-se, respectivamente, às ordens autoregressiva e de média móvel da componente sazonal.

Para se aplicar a metodologia de Box & Jenkins a série em estudo deve ser pelo menos estacionária de segunda ordem, ou seja, a série tem que ter média, variância e covariância finitas e constantes. O exame de estacionaridade pode ser realizado observando-se a Função de Autocorrelação ( $\rho_k$ ), FAC, da série em estudo. A referida função é estimada por:

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2} \quad (5)$$

onde  $T$  é o número de observações,  $\bar{y}$  é a média da série e  $k$  é o número de defasagens das autocorrelações (*lags*).

Se a série em estudo for não estacionária homogênea (Souza e Camargo, 1996) e com sazonalidade, temos um ARIMA Sazonal, ou seja, um SARIMA (p,d,q)(P,D,Q)s, onde  $d$  representa a ordem de diferenciação não sazonal e  $D$  a ordem de diferenciação sazonal necessárias para tornar a série estacionária. As diferenciações sazonal e não sazonal são realizadas da seguinte forma:

$$\begin{aligned} \text{diferença} &= (1-B)^d y_t & \text{diferença sazonal} &= (1-B^S)^D y_t \\ \text{para } d=1, (1-B)y_t &= y_t - y_{t-1} & \text{para } D=1, (1-B^S)y_t &= y_t - y_{t-DS} \end{aligned} \quad (6)$$

A identificação das ordens  $p$ ,  $q$ ,  $P$  e  $Q$  é realizada com base nas estimativas das Funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP), representados nos correlogramas (exposição gráfica das FAC e FACP contra as defasagens temporais). Para o caso de séries estacionárias não sazonais, a Tabela 1 fornece uma orientação para a identificação do modelo ARMA(p,q), com base nas características teóricas da FAC e da FACP.

**Tabela 1:** Padrões teóricos da FAC e da FACP

PROCESSO	FAC ( $\rho_k$ )	FACP ( $\phi_{kk}$ )
Ruído Branco	$\rho_k = 0 \forall k$	$\phi_{kk} = 1 \forall k$
AR(p)	$\rho_k$ decrescente	$\phi_{kk}$ nula para <i>lags</i> acima de $p$
MA(q)	$\rho_k$ nula para <i>lags</i> acima de $q$	$\phi_{kk}$ decrescente
ARMA(p,q)	$\rho_k$ idêntica ao de um AR(p) após $p$ - $q$ <i>lags</i>	$\phi_{kk}$ idêntica ao de um MA(q) após $p$ - $q$ <i>lags</i>

onde,  $\rho_k$  é a autocorrelação e  $\phi_{kk}$  é a autocorrelação parcial na defasagem de *lag*  $k$  das FAC e FACP, respectivamente.

Comparando-se as características teóricas da FAC e da FACP (Tabela 1) com as estimativas destas funções, obtidas com base na série em estudo (amostra), pode-se identificar o padrão teórico mais semelhante ao observado e assim determinar a ordem do modelo ARMA(p,q) que melhor explica a dinâmica da série em estudo. Na identificação da ordem da componente sazonal do modelo, deve-se verificar os valores amostrais da FAC e FACP separados por defasagens de comprimento  $s$ . Após a identificação da ordem do modelo, estimam-se os respectivos coeficientes  $\theta$ ,  $\Theta$ ,  $\phi$  e  $\Phi$  (Box e Jenkins, 1994).

A FAC e a FACP tem comportamentos teóricos conhecidos, entretanto na prática, como estas funções são estimadas, a análise dos gráficos da FAC e da FACP amostrais pode ser uma tarefa difícil, o que compromete a identificação precisa da ordem do modelo.

A incorporação de coeficientes adicionais (modelos de maior ordem) melhora o grau de ajustamento, reduzindo a soma dos quadrados dos resíduos. No entanto, segundo Box & Jenkins os modelos mais parcimoniosos produzem melhores previsões que os modelos sobreparametrizados.

Existem vários critérios de seleção de modelos que permitem encontrar um ponto de equilíbrio entre a redução na soma do quadrado dos resíduos e a parcimônia do modelo. Os critérios mais usados são o AIC (Akaike Information Criterion) e o BIC (Bayesian Information Criterion), cujas fórmulas são dadas por:

$$AIC = T \ln (\text{soma dos quadrados dos resíduos}) + 2n \quad (7)$$

$$BIC = T \ln (\text{soma dos quadrados dos resíduos}) + n \ln(T) \quad (8)$$

onde  $n$  é o número de parâmetros estimados e  $T$  é o número de observações da série temporal.

Comparando-se os valores de AIC e BIC de modelos com diferentes ordens, o melhor modelo é o que apresenta os menores valores nestas duas estatísticas.

No entanto, a seleção do melhor modelo não deve se basear apenas nos critérios AIC e BIC, a análise dos resíduos de modelos alternativos (concorrentes) ajustados é de extrema importância na escolha final do(s) modelo(s) que melhor explica(m) a dinâmica da série temporal em estudo. Se os resíduos são autocorrelacionados, então a dinâmica da série em estudo não é completamente explicada pelos coeficientes do modelo ajustado. Deve-se excluir do processo de escolha modelo(s) com esta característica. Uma análise da existência (ou não) da autocorrelação serial de resíduos é feita com base na estatística  $Q$  de Box-Pierce-Ljung, (Ljung e Box, 1978; Souza e Camargo, 1996) , representada formalmente como:

$$Q = T \cdot (T+2) \cdot \sum_{k=1}^K r_k^2 / (T-k), \quad (9)$$

$$r_k = \frac{\sum_{t=k+1}^T e_t e_{t-k}}{\sum_{t=1}^T e_t^2}, \quad (10)$$

onde  $r_k$  é a autocorrelação de ordem  $k$  dos resíduos  $e_t$  do modelo estimado.

A estatística  $Q$  é utilizada para testar se um conjunto de autocorrelações dos resíduos é (ou não) estatisticamente diferente de zero. Se os dados da série estudada são gerados por um processo estacionário, então a estatística  $Q \approx \chi_K^2$ . Observa-se que valores altos das autocorrelações dos resíduos implicam em valores altos de  $Q$ , enquanto que em um processo em que todas as autocorrelações de resíduos são nulas (ruído branco)  $Q$  é nulo. Se a estatística  $Q$  excede o valor tabelado de uma  $\chi_K^2$ , dado um certo nível de significância, rejeita-se a hipótese nula,  $H_0$ , de que os resíduos não são autocorrelacionados.

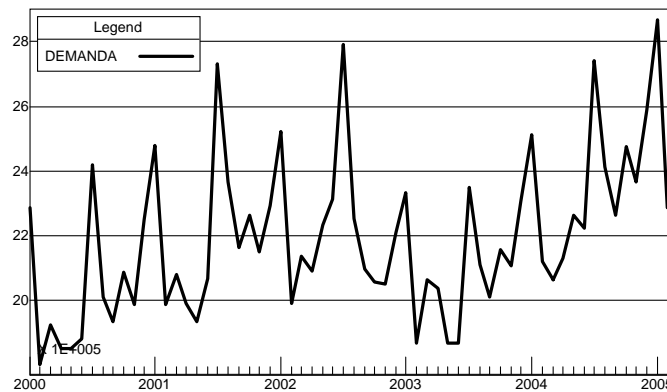
Um procedimento recomendado para identificar a melhor ordem de um modelo consiste em identificar um modelo inicial a partir da análise das estimativas da FAC e da FACP e em seguida fazer o teste da sobrefixação, onde são feitas várias análises para diferentes valores de  $p$ ,  $q$ ,  $P$ ,  $Q$  a partir do modelo inicial, em especial o teste t-student para verificar a significância de cada termo adicional na ordem do modelo, considerando-se as seguintes hipóteses:

$H_0$ : Coeficiente difere significativamente de 0

$H_1$ : Coeficiente não difere significativamente 0

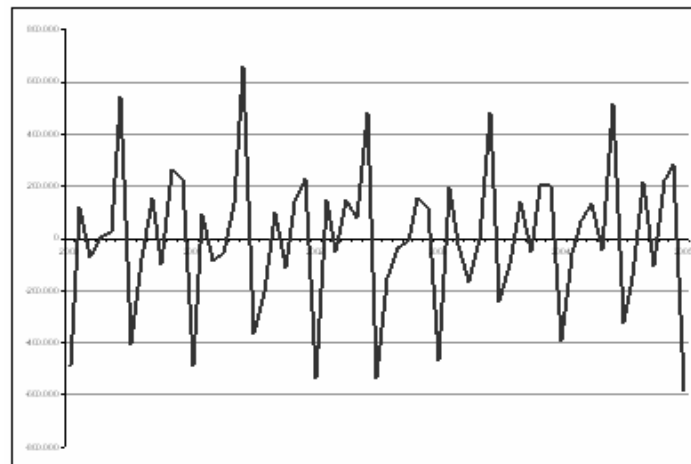
### 3. IDENTIFICAÇÃO E ESTIMAÇÃO DO MODELO

Para aplicar a metodologia de Box & Jenkins é necessário que a série tenha pelo menos estacionariedade de segunda ordem (média e variância constantes e autocovariâncias em função do intervalo de tempo entre as observações). A Figura 1 revela uma aparente tendência de crescimento da série passageiros-quilômetros transportados. Tal característica viola a hipótese de estacionariedade e por isso as observações precisam ser transformadas visando tornar a série estacionária.

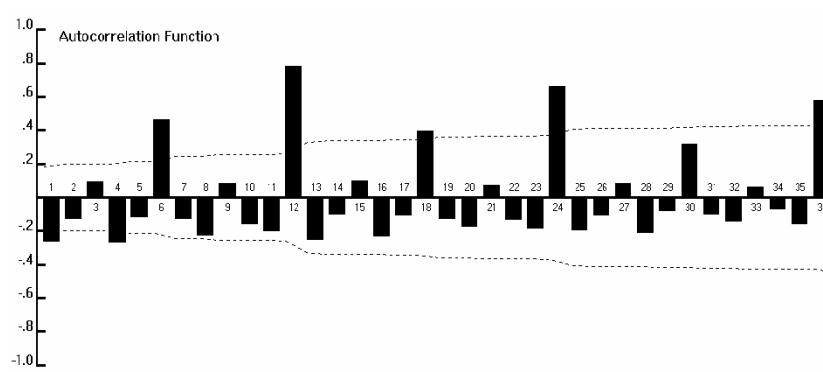


**Figura 1:** Utilização do transporte aéreo em vôos domésticos entre janeiro de 2000 e fevereiro de 2005

Para tornar a série estacionária na média foram tomadas as primeiras diferenças da série passageiros-quilômetros ( $d=1$ ), resultando na série apresentada na Figura 2, visualmente estacionária na média e na variância.



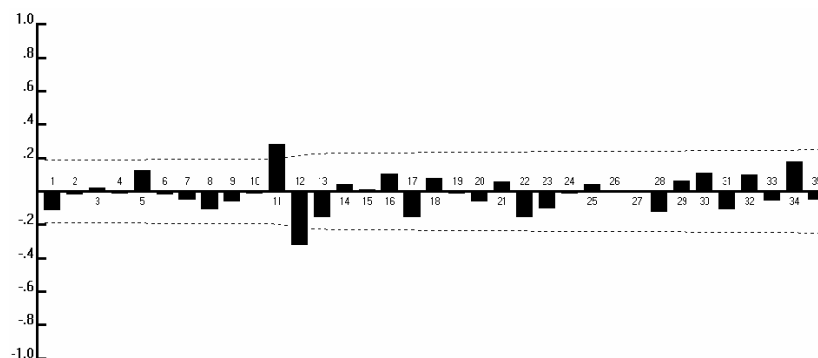
**Figura 2:** Série temporal após aplicação da primeira diferença



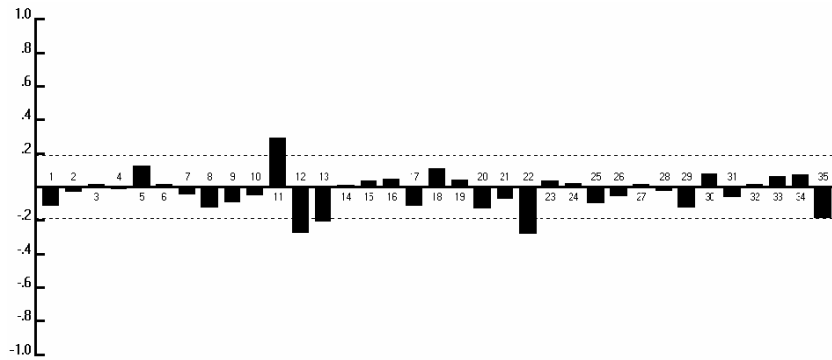
**Figura 3:** FAC amostral da série após aplicação da primeira diferença

A FAC da série diferenciada (Figura 3) sugere a necessidade de se fazer também uma diferença sazonal ( $D=1$ ) para tornar a série estacionária. Note que a autocorrelação em *lags* múltiplos de 12 decresce lentamente, uma característica de não estacionariedade.

A identificação da ordem do melhor modelo  $SARIMA(p,1,q) \times (P,1,Q)_{12}$ , gerador da série temporal, é realizada a partir da análise pictórica das estimativas das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da série transformada, apresentadas nas Figuras 4 e 5 respectivamente.

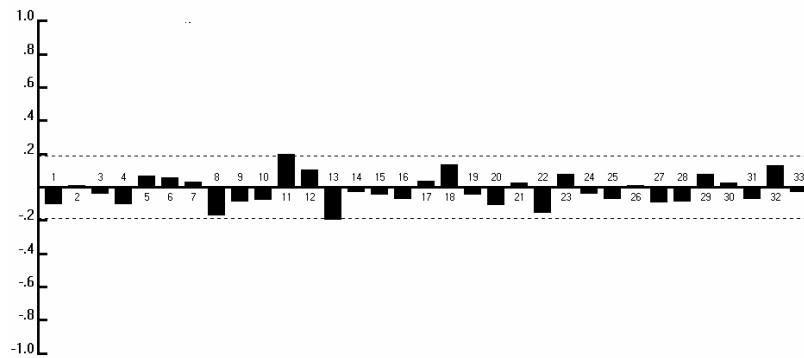


**Figura 4:** FAC amostral da série transformada



**Figura 5:** FACP amostral da série transformada

Assumindo um corte brusco no *lag* 12 (valores não significativos a partir do *lag* 12) do correlograma apresentado na Figura 4, pode-se considerar que na componente sazonal  $Q$  é 1 e  $P$  é 0. Quanto à ordem da componente não sazonal, ou seja,  $p$  e  $q$ , as Figuras 4 e 5 não exibem um comportamento típico de um AR ou MA, portanto no modelo inicial  $p$  e  $q$  são tomados como nulos. Resumindo o modelo inicial é um  $SARIMA(0,1,0) \times (0,1,1)_{12}$ , cujo coeficiente MA sazonal é  $\Phi_1 = 0,7898$ . A estimação foi baseada na série observada no período entre janeiro de 2000 e setembro de 2004. O coeficiente estimado é significativo e as estatísticas MAPE (*Mean Absolute Percentage Error*), AIC e RMSE (*Root Mean Square Error*) obtidas para este modelo são  $0,0248$ ;  $7,898 \times 10^4$  e  $7,678 \times 10^4$  respectivamente.



**Figura 6:** FAC dos resíduos do modelo  $SARIMA(0,1,0) \times (0,1,1)_{12}$

Na Figura 6, a FAC dos resíduos do modelo  $SARIMA(0,1,0) \times (0,1,1)_{12}$ , exibe um comportamento similar ao de um ruído branco, portanto, um MA na parte sazonal modela de forma satisfatória a série analisada.

Iniciando o teste de sobrefixação adiciona-se um termo AR na parte sazonal, resultando no modelo  $SARIMA(0,1,0) \times (1,1,1)_{12}$ . Conforme mostra a Tabela 2 este termo não é significativo.

**Tabela 2:** Adição de um termo AR na parte sazonal

Modelo	Coefficientes	Estatística t	MAPE	BIC	RMSE
$SARIMA(0,1,0) \times (1,1,1)_{12}$	$\Phi_1 = 0,1182$	1,0337 (0,696)	0,0365	$7,669 \times 10^4$	$9,1 \times 10^4$
	$\Theta_1 = 0,8112$	16,0483 (1,000)			

Em seguida adiciona-se mais um termo MA na parte sazonal, resultando no modelo  $SARIMA(0,1,0) \times (0,1,2)_{12}$ , e conforme mostra a Tabela 3 o segundo termo MA sazonal é não significativo.

**Tabela 3:** Adição de um termo MA na parte sazonal

Modelo	Coefficientes	Estatística t	MAPE	BIC	RMSE
$SARIMA(0,1,0) \times (0,1,2)_{12}$	$\Theta_1 = 0,0228$	7,0089 (1,000)	0,0361	$7,628 \times 10^4$	$9,07 \times 10^4$
	$\Theta_2 = 0,8696$	1,5533 (0,8764)			

Quanto a parte não sazonal do modelo, adiciona-se um termo MA, resultando no modelo  $SARIMA(0,1,1) \times (0,1,1)_{12}$ , cujas estatísticas são apresentadas na Tabela 4.

**Tabela 4:** Adição de um termo MA na parte não-sazonal

Modelo	Coefficientes	Estatística t	MAPE	BIC	RMSE
$SARIMA(0,1,1) \times (0,1,1)_{12}$	$\phi_1 = 0,0279$	0,1828 (0,1441)	0,0254	$8,24 \times 10^4$	$7,67 \times 10^4$
	$\Theta_1 = 0,8027$	12,9583 (1,000)			

Finalizando o teste de sobrefixação, um termo é adicionado na parte AR não sazonal do modelo resultando no SARIMA(1,1,0)x(0,1,1)<sub>12</sub> apresentado na Tabela 5.

**Tabela 5:** Adição de um termo AR na parte não-sazonal

Modelo	Coefficientes	Estatística t	MAPE	BIC	RMSE
SARIMA(1,1,0)x(0,1,1) <sub>12</sub>	$\phi_1 = -0,0132$ $\Theta_1 = 0,7851$	-0,0876(0,0694) 12,5358(1,000)	0,0252	8,24x10 <sup>4</sup>	7,709x10 <sup>4</sup>

As Tabelas 4 e 5 mostram que os termos AR e MA na parte não sazonal não são significativos ao nível de 5%. Portanto, o teste de sobrefixação confirma que para a série analisada, o melhor modelo é um SARIMA(0,1,0)x(0,1,1)<sub>12</sub>, cuja equação é:

$$(1-B^{12})(1-B)Y_t = (1-0,7898B^{12})a_t \quad (11)$$

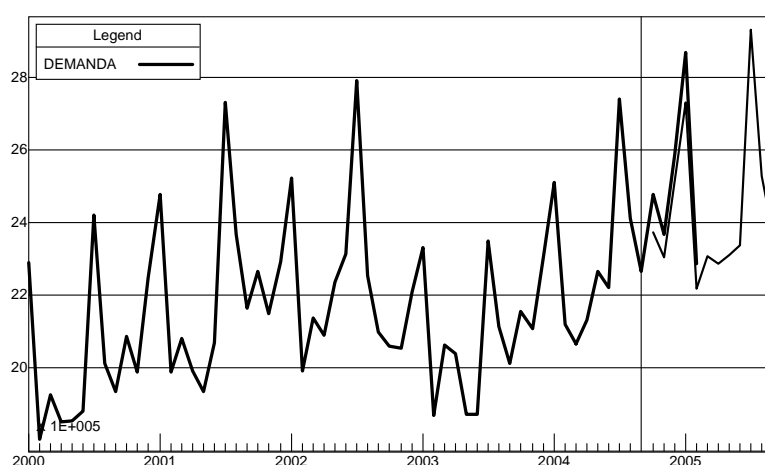
ou ainda

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + a_t - 0,7898 \cdot a_{t-12} \quad (12)$$

onde  $Y_t$  é o passageiros-quilômetros no mês  $t$ .

#### 4 ESTIMATIVAS E PREVISÕES

Neste item é apresentado o resultado alcançado com o modelos SARIMA(0,1,0)x(0,1,1)<sub>12</sub>. Na Figura 7 tem-se os valores ajustados e as previsões para o período de outubro/2004 a dezembro/2005.



**Figura 7:** Valores observados e previsões do modelo SARIMA(0,1,0)x(0,1,1)<sub>12</sub>

**Tabela 6:** Valores previstos e observados entre outubro de 2004 e fevereiro de 2005

Mês	Previsões SARIMA(0,1,0)x(0,1,1) <sub>12</sub>	Valor observado	Desvio (%)
Out/2004	2.372.196	2.476.409	-4,2
Nov/2004	2.303.900	2.367.836	-2,7
Dez/2004	2.515.984	2.588.184	-2,8
Jan/2005	2.732.646	2.869.290	-4,8
Fev/2005	2.216.601	2.287.128	-3,1

#### 5. CONCLUSÃO

Este trabalho teve como objetivo discutir e aplicar técnicas de previsão de séries temporais do tipo Box & Jenkins, para modelar o comportamento da demanda por transporte aéreo de passageiros em vôos domésticos, representada pela variável passageiros-quilômetros transportados, visando identificar um padrão de comportamento que permita realizar estimativas de valores futuros em um horizonte de curto prazo.

A partir de uma série com observações mensais no período janeiro de 2001 a dezembro de 2005, identificou-se um SARIMA(0,1,0)x(0,1,1)<sub>12</sub> como sendo o melhor modelo para representar a série da variável passageiros-quilômetros transportados.

Como sugestão para trabalhos futuros, recomenda-se investigar a aplicação de modelos estruturais (Durbin e Koopman, 2001) na série estudada, visando decompor a série em suas componentes não observáveis de tendência, sazonalidade e ciclo. Esta abordagem permite uma interpretação direta das componentes e oferece uma descrição dos traços estilizados da série, possibilitando analisar a mudança de tendência e alterações no padrão sazonal ou no comportamento cíclico da série. Uma característica que torna os modelos estruturais mais atraentes que os modelos Box e Jenkins é o fato deles relaxarem a restrição de estacionariedade da série analisada.

Os modelos estruturais também permitem um tratamento multivariado, onde séries de variáveis exógenas podem ser usadas para explicar a série de interesse. Esta característica é interessante, pois em geral, fatores macroeconômicos têm influência na utilização do transporte aéreo.

#### REFERÊNCIAS BIBLIOGRÁFICAS

- Box, G. E. P. and Jenkins, G. M e Reinsel, G. C. (1994) *Time Series Analysis: Forecasting and Control*. Prentice-Hall, Inc. USA.
- Durbin, J. e Koopman, S. J. (2001) *Time Series Analysis by State Space Methods*, Oxford.
- Goodrich, R. L. (1989) *Applied Statistical Forecasting* Business Forecast Systems INC. Belmont MA
- Instituto de Aviação Civil (2002) *Fluxo de Passageiros nas Ligações Aéreas Nacionais* (1ª ed.). Rio de Janeiro.
- Instituto de Aviação Civil (2001) *Demanda Global* (1ª ed). Rio de Janeiro.
- Ljung, G. M. e Box, G. E. P (1978) *On a measure of lack of fit in time series models*. Biometrika, Vol. 65, pp 297-303.
- Morettin, P. A. e Toloi, C. M. C. (1987) *Previsão de séries temporais* (2a ed.). Ed. Atual, São Paulo.
- Souza, R. C. e Camargo, M. E. (1996) *Análise e previsão de séries temporais : Os modelos ARIMA*, Ijuí, SEDGRAF.