

## 5

### PROPOSTAS DE MELHORIAS

Durante o estudo do trabalho realizado por Barcelos (2006), foram levantadas algumas questões que merecem ser abordadas na tentativa de tornar o processo de previsão mais preciso.

A primeira questão é referente ao método de ajuste dos modelos propostos aos dados históricos. O Método dos Mínimos Quadrados Não Lineares, explicado no Apêndice A, requer o fornecimento de um vetor de parâmetros inicial que no estudo de Barcelos corresponde ao vetor nulo, ou seja, os valores de todos os parâmetros do modelo são iguais a zero no primeiro momento. No entanto, em alguns casos, os modelos não aceitam como solução inicial o vetor nulo e Barcelos (2006) estima os valores sem critério pré-estabelecido, podendo com isso não ter encontrado a melhor solução possível para o ajuste.

A segunda questão corresponde ao método definido para efetuar a escolha do melhor modelo para cada série do estudo. Segundo Barcelos (2006), o modelo a ser escolhido para prever o futuro é o modelo que apresenta o melhor ajuste no passado, ou seja, o modelo que minimiza a diferença entre os valores ajustados e os valores reais no período de treinamento. No entanto, algumas séries apresentam comportamentos bem diferentes entre os períodos de treinamento e de validação, indicando que um bom ajuste aos dados passados não é garantia para uma boa previsão do futuro.

A terceira e última questão diz respeito à escolha de um bom modelo que independa do tamanho da série e da adição de novos dados históricos. Barcelos (2006) realiza o ajuste dos modelos para um único período, não garantindo assim que caso novos dados sejam adicionados à série histórica, o modelo escolhido continue sendo o melhor para efetuar previsões.

Nesse capítulo, são propostas metodologias que têm como objetivo tratar as questões listadas.

## 5.1

### ALGORITMO DE MELHORIA DO AJUSTE

Durante o desenvolvimento do trabalho, muitos dos modelos não aceitavam o valor nulo para seus parâmetros como solução inicial. Barcelos (2006), para esse problema, e atribui valores de maneira arbitrária até que algum ajuste seja efetuado. Esse questionamento fez surgir a dúvida se a solução inicial fornecida resultaria no melhor ajuste e se existe algum critério para definir os valores iniciais dos parâmetros.

Segundo Madsen et al. (2004), o método adequado para fornecer uma boa solução inicial é feito pelo entendimento detalhado do comportamento do modelo a ser ajustado e o cruzamento dessa análise com o comportamento dos dados históricos da série. Dada a existência de 72 modelos não lineares, a proposição de soluções iniciais se torna uma tarefa demorada, que pode não resultar em bons ajustes.

Dessa forma, o objetivo do algoritmo de melhoria do ajuste é definir diversas soluções iniciais para otimizar o ajuste dos modelos propostos aos dados da série histórica através da minimização da soma do quadrado dos resíduos (SQR) ou “*sum of squared residuals*” (SSR).

A Soma do Quadrado dos Resíduos (SQR) é uma medida da qualidade do ajuste que representa a variação dos valores ajustados com relação aos valores reais da série como é mostrado na eq. (42).

$$SQR = \sum_{i=1}^n (A_i - F_i)^2 \quad \text{Eq. (42)}$$

Onde:

- $n$  – número de períodos da série
- $A_i$  – valor real da série no período  $i$
- $F_i$  – valor do modelo para a série no período  $i$

Assim, o algoritmo busca melhorar o ajuste fornecendo diversas soluções iniciais para os modelos mantendo a solução que minimiza o SQR como é mostrado na Figura 14.

**Parte 1 – Gerando as Soluções Iniciais**

```

1  Table Modelos
2  Smpl {!inicioajuste} {!fimajuste}
3  For !i=1 to 72
4      %x=Modelos(!i,1)
5      %y=Modelos(!i,2)
6      c=0
7      Equation {%x}.ls {%y}
8  Next

```

**Parte 2 – Otimizando o Ajuste**

```

9  For !i=1 to 72
10     %x=Modelos(!i,1)
11     %y=Modelos(!i,2)
12     !old=@errorcount
13     !SQR={%x}.@ssr
14     !new=@errorcount
15     If !new>!old Then
16         !SQR=M
17     End If
18     For !j=1 to 72
19         %w=Modelos(!j,1)
20         {%w}.updatecoefs
21         Equation teste.ls {%y}
22         !SQRnovo=teste.@ssr
23         If !SQRnovo<!SQR Then
24             !SQR=!SQRnovo
25             {%w}.updatecoefs
26             Equation {%x}.ls {%y}
27         End If
28     Next
29     Modelos(!i,3)=!SQR
30 Next

```

Figura 14 – Algoritmo de Melhoria do Ajuste

Fonte: Autor

O algoritmo desenvolvido parte do princípio que os valores dos parâmetros obtidos no ajuste de um modelo podem ser utilizados como os valores iniciais para os parâmetros em outro modelo.

Dessa forma, a primeira parte consiste em gerar as primeiras soluções considerando o vetor nulo como sendo o valor dos parâmetros do modelo, como foi feito no trabalho de Barcelos (2006).

A matriz “Modelos” é criada (linha 1) contendo a identificação dos modelos na primeira coluna e a equação dos modelos na segunda coluna. Em seguida, antes de iniciar o processo recursivo de estimação (linha 3 / linha 8), é necessário determinar o período do ajuste utilizando a função “smp1” (linha 2) que requer o ano inicial e o ano final associados as variáveis numéricas “!inicioajuste” e “!fimajuste”.

Através da variável numérica “i” que percorre as linhas da tabela, as variáveis textuais “x” e “y” recebem a identificação (linha4) e equação (linha 5) do modelo respectivamente. Assumindo que os valores dos parâmetros correspondem à um vetor nulo “c” (linha 6), a função “Equation” é chamada (linha 7) para realizar o ajuste, sendo necessário informar o nome do modelo, a equação e o método de estimação. A representação do Método dos Mínimos Quadrados Não Lineares é feita pelo comando “*ls – Least Squares*”.

Ao final da primeira parte, cada modelo que conseguiu realizar o ajuste apresenta uma tabela de informações, como é exemplificado na Figura 15.

The screenshot shows a software window titled "Equation: BARCELOSIX9" with a menu bar (View, Proc, Object, Print, Name, Freeze, Estimate, Forecast, Stats, Resids). The main area displays the following information:

Dependent Variable: R3  
 Method: Least Squares  
 Date: 10/30/12 Time: 08:16  
 Sample (adjusted): 1965 2002  
 Included observations: 38 after adjustments  
 Convergence achieved after 107 iterations  

$$R3=R3(-1)+(C(1)*EXP(-C(2)*R3(-1)+C(3)*R3(-1)^C(4)))-((C(5)/R3(-1))*EXP(C(3)*R3(-1)^C(4)))$$

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-337.9724	160.4519	-2.106378	0.0429
C(2)	0.007576	0.000234	32.35581	0.0000
C(3)	-31845.20	83839.41	-0.379836	0.7065
C(4)	-2.564424	0.755940	-3.392365	0.0018
C(5)	-16544.12	7584.800	-2.181221	0.0364

R-squared	0.992128	Mean dependent var	128.5486
Adjusted R-squared	0.991173	S.D. dependent var	29.87281
S.E. of regression	2.806543	Akaike info criterion	5.023863
Sum squared resid	259.9306	Schwarz criterion	5.239335
Log likelihood	-90.45340	Hannan-Quinn criter.	5.100527
Durbin-Watson stat	1.342041		

Figura 15 – Resumo de Informações do Ajuste

Fonte: Autor

Essa tabela contém o SQR do ajuste e os novos valores encontrados para os parâmetros, informações necessárias para o início do processo de otimização do algoritmo.

O processo de otimização, segunda parte do algoritmo, tem início verificando o SQR atual do modelo “i” para, posteriormente, comparar esse valor com o SQR dos novos ajustes do mesmo modelo. Essa atividade é realizada utilizando a função “@ssr” que lê as informações do modelo identificado pela variável textual “x” (linha 10) e aloca o valor do SQR atual à variável numérica “SQR” (linha 13). No entanto, o valor do SQR só pode ser alocado caso o modelo tenha realizado com sucesso o ajuste na primeira fase.

Essa verificação é feita utilizando a função “@errorcount” que aloca nas variáveis “old” (linha 12) e “new” (linha 14) o número de erros do algoritmo antes e depois da tentativa de ler o SQR. Caso o número de erros aumente (linha 15), representado o insucesso do modelo em realizar o ajuste com o vetor nulo na primeira parte, o algoritmo associa à variável numérica “SQR” um valor “M” muito grande (linha 16).

Após calcular o SQR, o próximo passo consiste em realizar novos ajustes do modelo “i” considerando como solução inicial os valores dos parâmetros encontrados nos ajustes realizados na primeira parte do algoritmo para os modelos “j”. Essa etapa se inicia através da função “updatecoefs” (linha 20) que atualiza o vetor “c” de soluções iniciais com os valores dos parâmetros do modelo “j” associado a variável textual “w” (linha 19). Em seguida, utilizando a equação do modelo “i” associada a variável textual “y” (linha 11), efetua-se um novo ajuste nomeado “teste” (linha 21) e aloca-se o SQR desse ajuste à variável numérica “SQRnovo” (linha 22). Caso o SQR do novo ajuste seja menor que o SQR obtido na primeira parte (linha 23), as soluções do modelo “j” são novamente adotadas (linha 25) e o modelo “i” é novamente ajustado com sua própria identificação (linha 26) substituindo todos os dados da tabela de informações do ajuste.

Após testar todas as soluções dos modelos “j” como soluções iniciais para o modelo “i”, o algoritmo registra o SQR final na terceira coluna da tabela de “Modelos” (linha 29) para facilitar o procedimento de escolha do melhor modelo que é feito posteriormente.

## 5.2

### METODOLOGIA DE ESCOLHA DO MODELO

Uma das maiores preocupações que pode surgir durante o processo de previsão é se o modelo proposto continuará a realizar boas previsões à medida que novos dados históricos são acrescentados à série. Barcelos (2006), em seu trabalho, não levanta essa questão e realiza o ajuste dos modelos para um único período.

As curvas de volume recuperável, de uma maneira geral, tendem a possuir um comportamento em forma de S como é mostrado na Figura 16.

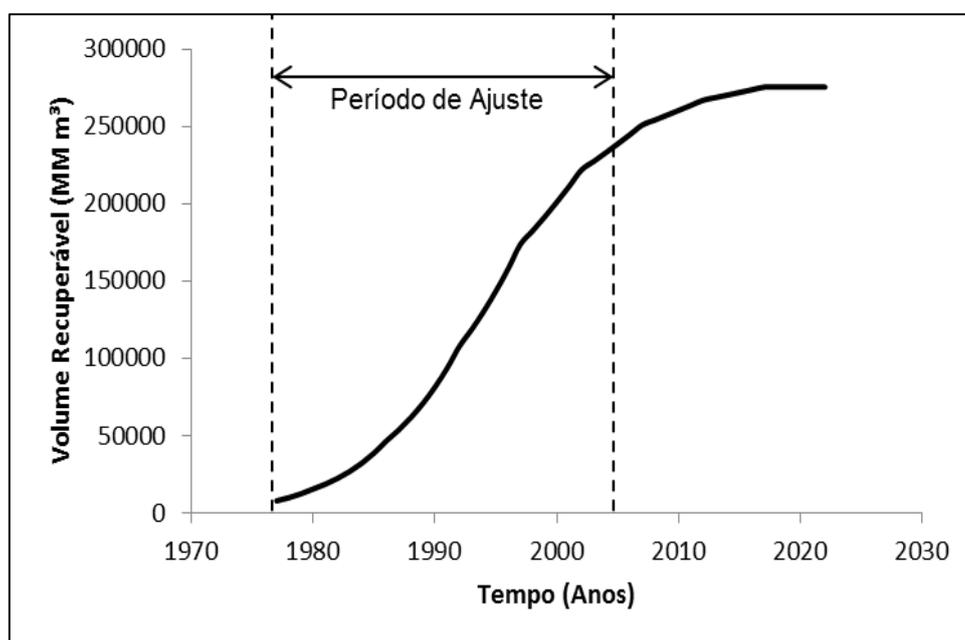


Figura 16 – Análise de Tendência de Crescimento da Série

Fonte: Autor

Caso exista uma tendência de crescimento acelerada na maior parte do período de ajuste, o modelo tentará manter o comportamento de crescimento em suas previsões, podendo diferir dos valores reais da série que começam a desacelerar.

Levando em conta essas considerações, a metodologia de escolha tem como objetivo verificar se existe algum modelo que realize o melhor ajuste dentre todos os modelos, independente da incorporação de novos dados à série histórica. No entanto, caso não seja possível, o modelo escolhido deve ser o que melhor representa o atual momento da série.

Diferente do algoritmo de otimização, que é totalmente realizado de maneira automática pelo E-views, a metodologia de escolha do modelo envolve, em sua maior parte, uma análise manual realizada pelo analista responsável pelas previsões como é mostrado na Figura 17.

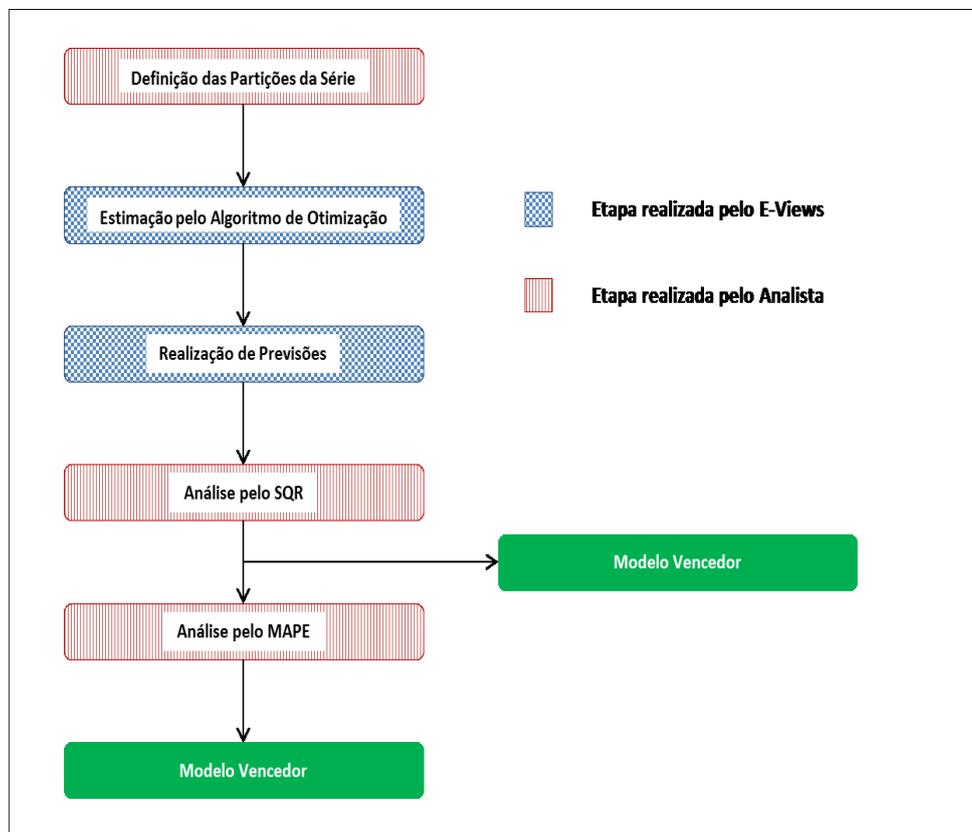


Figura 17 – Etapas da Metodologia de Escolha do Modelo

Fonte: Autor

A primeira etapa consiste em definir as partições da série original, ou seja, dividir a série histórica de dados em um período de treinamento, sobre o qual é feito o ajuste, e um período de validação para avaliar a precisão dos valores previstos.

Como foi mencionado no capítulo 4, cada série histórica sofreu um total de quatro partições, sendo elas:

- Partição 1 – Ajuste: Ano Inicial – 2002 / Validação: 2003 – 2007
- Partição 2 – Ajuste: Ano Inicial – 2003 / Validação: 2004 – 2008
- Partição 3 – Ajuste: Ano Inicial – 2004 / Validação: 2005 – 2009
- Partição 4 – Ajuste: Ano Inicial – 2005 / Validação: 2006 – 2010

A criação de diversas partições tem como objetivo simular a incorporação de novos dados à série histórica, fato que ocorre todos os anos, e verificar se um mesmo modelo gera o melhor ajuste sempre.

A segunda e terceira etapa do processo consistem em aplicar o algoritmo de otimização da estimação e realizar as previsões de todos os modelos para cada uma das partições. O processo de previsão também é um algoritmo realizado pelo software E-Views como é mostrado na Figura 18.

1	<b>Smpl</b> {!inicioprevisao} {!fimprevisao}
2	<b>For</b> !i=1 <b>to</b> 72
3	%x=Modelos(!i,1)
4	%y=Modelos(!i,4)
5	{%x}.Forecast {%y}
6	<b>Next</b>

Figura 18 – Algoritmo de Previsão

Fonte: Autor

Esse algoritmo é sempre executado após o algoritmo de otimização da estimação e se inicia pela determinação do período de previsão utilizando a função “smpl” (linha 1) que requer o ano inicial e o ano final associados as variáveis numéricas “!inicioprevisao” e “!fimprevisao”.

Em seguida, através de um processo recursivo (linha 2 / linha 6), cada modelo representado pela variável textual “x” é chamado pela função “forecast” que efetua a previsão para o período determinado. A identificação do modelo deve ser a mesma utilizada durante o ajuste e um novo nome simplificado, associado a variável “y”, deve ser fornecido para que a série de previsão seja armazenada.

Ao final da segunda e terceira etapa, todos os modelos que não conseguiram efetuar o ajuste ou a previsão em qualquer uma das partições devem ser eliminados e desconsiderados do processo de escolha, que tem sua continuidade através de uma análise do SQR.

Nessa etapa, uma tabela contendo o SQR de todos os modelos para todas as partições é montada e os mesmos são ordenados do menor para o maior indicando qual o modelo que obteve o melhor ajuste em cada partição, como é exemplificado na Tabela 1.

Período de Ajuste							
1974 - 2002		1974 - 2003		1974 - 2004		1974 - 2005	
Modelos	SQR	Modelos	SQR	Modelos	SQR	Modelos	SQR
Barcelos7x1	33.788	Barcelos7x1	34.879	Barcelos7x1	34.893	Barcelos7x1	34.894
Barcelos7x8	33.962	Barcelos7x2	46.571	Barcelos7x8	34.922	Barcelos7x8	42.992
Barcelos7x9	41.733	Barcelos7x5	46.596	Barcelos7x9	37.826	Barcelos7x2	48.744
Barcelos7x5	41.833	Barcelos7x6	48.705	Barcelos7x2	47.560	Barcelos7x5	48.769
Barcelos7x2	41.924	Barcelos1x8	54.042	Barcelos7x5	47.605	Barcelos7x6	50.878

Tabela 1 – Análise do SQR

Fazendo uma análise sobre os resultados da tabela, deve-se verificar se todas as partições apresentam o mesmo modelo com o melhor ajuste. Caso isso ocorra, tem-se um modelo que não é afetado pela adição de novos dados a série histórica e pode ser utilizado para representar o passado e fazer estimativas do futuro encerrando o processo de escolha.

No entanto, caso cada partição apresente um modelo vencedor diferente, deve-se aceitar que a série histórica é uma série instável e não pode ser representada por um único modelo. Sendo assim, deve-se obter um modelo que melhor represente o momento atual da série.

Essa análise é realizada através da comparação dos valores previstos no período de validação com os valores reais, e é feita utilizando o conceito do Erro Percentual Absoluto Médio (MAPE – Mean Absolute Percentage Error) mostrado na eq. (43).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|A_i - F_i|}{A_i} \quad \text{Eq. (43)}$$

Onde:

- $n$  – número de observações
- $A_i$  – valor real para a observação  $i$
- $F_i$  – valor previsto para a observação  $i$

O MAPE indica a diferença percentual média entre os valores previstos e os valores reais e por isso, quanto menor o valor do MAPE melhor é a previsão. Assim, ao calcular o MAPE para cada modelo nas 20 observações (5 previsões para cada uma das 4 partições), pode-se construir uma nova tabela ordenando os

valores de maneira crescente e assim determinar o modelo vencedor, como é exemplificado na Tabela 2.

<b>Modelos</b>	<b>MAPE</b>
Barcelos2x3	3,19%
Barcelos7x1	3,74%
Barcelos4x9	4,33%
Barcelos5x3	4,60%
Barcelos4x8	6,27%

Tabela 2 – Análise pelo MAPE