

WAVELETS E PREVISÕES DE SÉRIES DE TEMPO: UMA ANÁLISE EMPÍRICA

Guilherme V. Homsy¹
Marcelo S. Portugal²
Jorge P. Araújo³

Resumo: O presente artigo compara, sob a forma de estudo de casos, previsões relativas à três diferentes métodos de modelagem de séries de tempo, os quais consistem na aplicação da metodologia ARIMA tanto da forma usual quanto amparada por dois procedimentos auxiliares baseados em análise de ondaletas (*wavelets*) para três séries de tempo econômicas. O primeiro procedimento consiste em fracionar cada uma das séries em duas sub-séries, tendo em vista a análise tradicional de maneira separada para posterior fusão das previsões. O segundo procedimento consiste no emprego da análise tradicional mediante alisamento prévio das séries.

1. Introdução

A análise de ondaletas (*wavelets*) tem se mostrado, a partir de meados dos anos 80, uma ferramenta de bastante utilidade em vários campos do conhecimento como física, matemática, engenharia e, na presente década, em economia podendo-se ressaltar, em econometria, sua utilização como procedimento auxiliar para ajuste e previsão de séries de tempo.

No sentido em questão, pode-se ressaltar duas contribuições de grande relevância:

- Donoho e Johnstone (1992a,1992b,1993), Donoho et al. (1993,1995) e Johnstone e Silverman (1996) propõem um procedimento para se obter, à partir de séries *originais* de tempo, séries *alisadas* (*denoised*), que são possíveis pontos de partida para a aplicação de metodologias de previsão;
- Ariño e Vidakovic (1995) e Ariño (1995) propõem um procedimento alternativo (AV) à modelagem de uma dada série de tempo $\{y_t\}_{t=0}^{T-1}$ através de uma metodologia principal (como, por exemplo, ARIMA), o qual consiste em utilizar a análise de ondaletas para fracionar a série original em duas sub-séries, sendo uma relativa às “altas” frequências e a outra às “baixas” frequências. Posteriormente, aplica-se uma determinada metodologia para se proceder a modelagem, de forma separada, em cada uma dessas sub-séries, tendo como base melhor detectabilidade do comportamento cíclico da série em questão.

O presente artigo tem como objetivo aplicar, de forma isolada, cada um dos procedimentos em questão com relação à metodologia ARIMA e comparar, ainda que de maneira desprovida de poder estatístico, as previsões obtidas com as relativas à aplicação “tradicional” da mesma para três séries de tempo. O presente artigo é o início de uma pesquisa que visa estender a utilização da análise de ondaletas para o estudo de séries de tempo com o objetivo principal de tornar possível sua futura utilização como meio principal (não mais auxiliar) de previsão.

2. Ondaletas

2.1. Definição

Ondaletas podem ser consideradas, com relação ao presente estudo, como sendo funções da forma $w_{j,k}(t)$, geradas à partir de translações e dilatações de uma única função $w(t)$, que

¹ Doutorando em Economia da UFRGS.

² Professor do Programa de Pós-Graduação da UFRGS e Pesquisador do CNPq.

³ Professor do Programa de Pós-Graduação da UFRGS.

permitem a representação, tanto no domínio das abscissas quanto no da frequência, dos elementos de determinados subespaços do conjunto das funções (F).

Com relação à análise em questão, a variável independente t representa o tempo, e seu domínio é referido como *domínio do tempo*.

2.2 Histórico

A possibilidade de representar (ou codificar) funções através de coeficientes relacionados a determinados sistemas de funções sempre despertou profundo interesse, devendo-se à mesma inúmeras aplicações nas mais variadas áreas do conhecimento.

Em 1807, Fourier afirmou que qualquer função $f(t)$ definida no intervalo $[-p, p]$ poderia ser expressa através de uma série trigonométrica da forma $f(t) = \frac{a_0}{2} + \sum_{n=1}^{+\infty} a_n \cdot \cos(nt) + b_n \cdot \sin(nt)$.

Combinando-se tal afirmação com as noções atuais de função, integral e convergência, resulta que o conjunto de funções $\{\sin(nt), \cos(nt)\}_{n \in \mathbb{N}}$ constituiria um *conjunto gerador* das funções definidas no intervalo $[-p, p]$ o qual, munido do produto interno

$$\langle u, v \rangle = \int_{-p}^p u(t) \cdot \overline{v(t)} dt \quad (2.1)$$

, formaria uma base ortogonal para tais funções, sendo os coeficientes associados à uma determinada função de acordo com a mesma dados por

$$\begin{cases} a_0 = \frac{\langle f(t), 1 \rangle}{\langle 1, 1 \rangle} = \frac{1}{p} \int_{-p}^p f(t) dt, & a_n = \frac{\langle f(t), \cos(nt) \rangle}{\langle \cos(nt), \cos(nt) \rangle} = \frac{1}{p} \int_{-p}^p f(t) \cos(nt) dt \\ b_n = \frac{\langle f(t), \sin(nt) \rangle}{\langle \sin(nt), \sin(nt) \rangle} = \frac{1}{p} \int_{-p}^p f(t) \sin(nt) dt \end{cases} \quad (2.2).$$

Considerando-se a relação de Euler, poder-se-ia rescrever tal representação de acordo com a forma

$$f(t) = \sum_{n=0}^{+\infty} c_n \cdot e^{int} \quad (2.3) \quad , \text{ onde} \quad c_n = \frac{1}{p} \int_{-p}^p f(t) e^{-int} dt \quad (2.4)$$

são denominadas *transformadas de Fourier de $f(t)$* .

Ademais, pode-se substituir a representação acima por:

$$f(t) = \sum_{n=0}^{+\infty} c_n \cdot \mathbf{y}_n(t) \quad (2.5) \quad \text{onde} \quad \begin{cases} \mathbf{y}(x) = e^{it} & (2.6) \\ \mathbf{y}_n(x) = \mathbf{y}(nt) & (2.7) \end{cases}.$$

Note-se que o sistema de equações (2.5)-(2.7) sugere a possibilidade de escrever qualquer função $f(t)$ definida em $[-p, p]$ como sendo uma soma infinita entre funções com diferentes períodos (e, conseqüentemente, frequências), geradas através de compressões da função $\mathbf{y}(t)$.

Contudo, em 1873, Paul Du Bois-Reymont apresentou uma função contínua e periódica de período $2p$ cuja série de Fourier, como passaram a ser chamadas as séries acima, divergia num

determinado ponto. Posteriormente, outros exemplos surgiram, como o exemplo de Fejér, em 1910, de uma função contínua cuja série de Fourier divergia em todos os pontos.

Ademais, os coeficientes obtidos de acordo com a base em questão, apesar de representarem $f(t)$ de acordo com diferentes níveis de frequências, a dissociam de representação no próprio eixo do tempo.

Tendo em vista as limitações apresentadas pela análise de Fourier, três direções surgiram do trabalho de Du Bois-Reymont:

- i) modificar a definição de função, de forma a encontrar uma que se adaptasse às séries de Fourier;
- ii) modificar a noção de convergência de forma a restabelecer a igualdade entre a função e a série;
- iii) descobrir outros conjuntos de funções, que não $\cos(nt)$ e $\sin(nt)$, com os quais se pudesse gerar séries que convergem para a função.

Em 1909, Haar determinou um sistema ortonormal $\{h_n(t)\}$, $n \in \mathbb{N}$, gerado através de rotações e translações de uma função $h_0(t)$ tal que, dada $f(t) \in C^0[0,1]$, em relação ao produto interno (2.1), a representação da mesma no sistema proposto, ou seja,

$$\langle f, h_0 \rangle \cdot h_0 + \langle f, h_1 \rangle \cdot h_1 + \dots + \langle f, h_n \rangle \cdot h_n + \dots \quad (2.8)$$

converge uniformemente para $f(t)$. Tal sistema, denominado *sistema de Haar*, possuía a vantagem de ser composto por funções de suporte compacto (e, portanto, capaz de localizar sinais tanto no domínio da frequência como o sistema trigonométrico quanto no domínio do tempo). O trabalho de Haar continuou nas décadas de 20 e 30 com Schauder e Lusin.

A base de Haar tem o inconveniente de ser formado por funções descontínuas e ser inapropriado para funções de classe C^1 . Schauder apresentou uma base, $\Delta_1, \Delta_2, \dots$ de funções contínuas, que não são de classe C^1 . Dada $f(t)$ contínua, ele mostrou que

$$f(t) = a + b + \sum_1^{+\infty} a_n \cdot \Delta_n(t) \quad (2.9)$$

e que a convergência é uniforme em $[0,1]$. A base de Schauder apresenta vantagens sobre a de Haar, pois se $f(t)$ é de classe C^1 , derivando a série obtém-se uma série para a derivada $f'(t)$ em termos de base de Haar, cuja convergência é uniforme.

Posteriormente, o foco da análise em questão passou a abranger, além da determinação de bases para os espaços em estudo, metodologias de obtenção das mesmas a partir de funções ditas *originais* (como, por exemplo, a função $h_0(t)$ de Haar) através de algum processo de associação sendo que, a partir disto, surgem três pontos importantes a serem ressaltados:

- Tendo-se em vista o objetivo de se representar funções tanto no domínio do tempo quanto da frequência, os processos de associação em questão deveriam considerar não apenas compressões/dilatações de funções originais como ocorria no caso da análise de Fourier, mas também translações das mesmas. Dessa forma, em termos de representação matemática, funções não-originais passariam a ser representadas a partir das originais através do uso de dois índices, e não apenas de um como no caso da fórmula (2.5);

- Tendo-se em vista o objetivo da inclusão da representação temporal, certas condições devem ser impostas acerca de uma dada função para que a mesma possa ser considerada originária de um dado sistema dentre as quais não figura, ao contrário de uma eventual impressão inicial, o caráter de suporte compacto. De qualquer forma, a mesma não pode apresentar um caráter análogo, por exemplo, ao oscilatório apresentado pelas funções originais relativas à análise de Fourier;

• Como uma questão de nomenclatura, convencionou-se denominar as funções que atendessem às condições supra-referidas com o termo *ondaletas*. Devido ao caráter de reversibilidade das operações de rotação e translação de funções, tal denominação torna-se extensiva a qualquer elemento do sistema obtido a partir da aplicação de tais operações às mesmas. Sendo assim, em um determinado sistema formado por ondaletas para um determinado espaço, a determinação da função original torna-se uma questão de caráter arbitrário.

Levando-se em conta tais considerações, Grossmann e Morlet propuseram que, dada uma função original $\mathbf{w}(t)$, se procedesse às rotações e translações de acordo com a forma

$$\mathbf{w}_{a,b}(t) = |a|^{-\frac{1}{2}} \mathbf{w}\left(\frac{t-b}{a}\right), \text{ onde } \begin{cases} a \in \mathfrak{R}^+ \\ b \in \mathfrak{R} \end{cases} \quad (2.10)$$

obtendo-se portanto o sistema $\left\{ \mathbf{w}_{a,b}(t) \right\}_{\substack{a \in \mathfrak{R}^+ \\ b \in \mathfrak{R}}}$. A função original $\mathbf{w}(t)$ corresponde ao caso em que $a=1$ e $b=0$.

O trabalho de Haar foi completado em 1988 por Ingrid Daubechies em seu artigo "*Orthonormal Bases of compactly supported wavelets*", a partir do qual resulta a análise de ondaletas relativa ao presente artigo. Através da aplicação de translações diádicas e dilatações binárias à funções originais indexadas por números naturais constrói-se um conjunto de bases ortonormais de suporte compacto para o $L^2(\mathfrak{R})$ no sentido de média quadrática.

2.3. Considerações acerca da análise de ondaletas

2.3.1 Translações Diádicas e Dilatações Binárias

O procedimento de translações diádicas e dilatações binárias consiste em representar tais operações aplicadas à função original $\mathbf{w}(t)$ de acordo com a seguinte forma (escolhida para as demais seções do presente estudo):

$$\mathbf{w}_{j,k}(t) = 2^{-\frac{j}{2}} \mathbf{w}(2^{-j}t - k), \quad j, k \in \mathbb{Z} \quad (2.11) \quad \text{ou} \quad \mathbf{w}_{j,k}(t) = 2^{\frac{j}{2}} \mathbf{w}(2^j t - k), \quad j, k \in \mathbb{Z} \quad (2.12)$$

ou seja, em formar o sistema através de funções cujas frequências (e, logicamente, medida dos suportes, no caso da função original possuí-lo compacto) associadas ao índice j relacionem-se com os referentes à mesma de acordo com exponenciações inteiras na base dois.

Cabe ressaltar que, no caso de $\mathbf{w}(t)$ possuir norma unitária e suporte compacto, todos os elementos do sistema formado apresentarão tal propriedade. Para um determinado elemento genérico $\mathbf{w}_{j,k}(t)$, ter-se-á

$$\langle \mathbf{w}_{j,k}(t), \mathbf{w}_{j,k}(t) \rangle = \int_{-\infty}^{+\infty} |\mathbf{w}_{j,k}(t)|^2 dt = \left(2^{\frac{j}{2}} \right)^2 \int_{-\infty}^{+\infty} |\mathbf{w}(2^j t - k)|^2 dt = 2^j \cdot \frac{1}{2^j} \int_{-\infty}^{+\infty} |\mathbf{w}(t)|^2 dt = \langle \mathbf{w}(t), \mathbf{w}(t) \rangle = 1 \quad (2.13)$$

de forma que, se o sistema $\left\{ \mathbf{w}_{j,k} \right\}_{j,k \in \mathbb{Z}}$ consistir base para algum determinado subespaço de F , o mesmo será ortonormal.

É importante ressaltar que, tendo em vista a propriedade de representação em tempo-frequência, variações no índice j relacionam-se às medidas dos intervalos nos quais as ondaletas concentram maior parte de sua oscilação, de forma a se referir a tal índice como sendo relativo ao

nível de resolução associado às ondaletas em estudo. Enquanto que na representação (2.12) valores maiores do mesmo referem-se à ondaletas de alta resolução, adequados à caracterização de fenômenos efêmeros associados à função $f(t)$, ocorre o oposto com relação à representação (2.11)

Considerar-se-ão, na seqüência de análise, funções originais que apresentem suporte compacto. No caso de o sistema em questão consistir base para o espaço de interesse $(L^2(\mathfrak{R}))$ então, para qualquer função $f(t)$ pertencente à tal espaço, existirão $\{b_{j,k}\}_{j,k \in Z}$ tais que a soma

$$\sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} b_{j,k} \mathbf{w}_{j,k}(t) \quad (2.14)$$

convergirá em média quadrática para $f(t)$. Denomina-se os coeficientes $b_{j,k}$, em analogia aos coeficientes c_n relativos à análise de Fourier, *coeficientes de ondaletas* de $f(t)$.

2.3.2. Classificação Genealógica de Sistemas de Ondaletas

Considerar-se-ão dois tipos de sistemas de ondaletas, a serem classificados tendo como base tanto as propriedades apresentadas pelas funções $\mathbf{w}_{j,k}(t)$ pertencentes aos mesmos quanto as inter-relações destas com as relativas ao sistema de classificação diferente.

Sistema I $\{\mathbf{y}_{j,k}\}_{j,k \in Z}$

Tal sistema, base ortonormal de $L^2(\mathfrak{R})$, é gerado à partir da aplicação de translações diádicas e dilatações binárias à uma função original $\mathbf{y}(t)$ (denominada ondaleta-mãe), que atende às seguintes condições:

$$\bullet \forall s \in \{0, \dots, r-1\}, r \in \mathbb{N}, \int_{-\infty}^{+\infty} t^s \mathbf{y}(t) dt = 0 \quad (2.15);$$

$$\bullet \int_{-\infty}^{+\infty} t^r \mathbf{y}(t) dt < +\infty \quad (2.16).$$

Para um dado valor do índice j , costuma-se denotar como W_j o conjunto gerado à partir do sistema $\{\mathbf{y}_{j,k}\}_{k \in Z}$ sendo que, à partir da suposição de que o sistema $\{\mathbf{y}_{j,k}\}_{j,k \in Z}$ constitua base para o espaço $L^2(\mathfrak{R})$, conclui-se que $\bigcup_{k \in Z} W_j = L^2(\mathfrak{R})$.

Sistema II $\{\mathbf{f}_{j,k}\}_{j,k \in Z}$

Tal sistema normal, linearmente dependente e gerador de $L^2(\mathfrak{R})$, é obtido à partir da aplicação de translações diádicas e dilatações binárias à uma função original $\mathbf{f}(t)$ (denominada ondaleta-pai), que atende às seguintes condições:

$$\bullet \forall j \in Z, \exists \{l_k\}_{k \in Z} \quad \text{tal que} \quad \mathbf{f}_{j,k}(t) = \sum_{k \in Z} l_k \mathbf{f}_{j-1,k} \quad (2.17)$$

, ou seja, toda ondaleta relativa ao nível de resolução j pode ser escrita como combinação linear das ondaletas referentes ao nível de resolução imediatamente inferior e, por conseqüência, a todos os níveis de resolução inferiores;

$$\bullet \forall j, j', k, k' \in Z \quad , \quad \langle \mathbf{y}_{j,k}, \mathbf{f}_{j',k'} \rangle = 0, \quad j' \geq j \quad (2.18)$$

, do que se pode concluir que cada ondaleta referente a este sistema é ortogonal às ondaletas de nível de resolução inferior referentes ao sistema anterior;

$$\bullet \forall k, k' \in Z \quad , \quad \langle \mathbf{f}_{j,k}, \mathbf{f}_{j,k'} \rangle = 0 \quad (2.19)$$

, isto é, ondaletas referentes a um dado nível de resolução j formam um conjunto ortonormal, cujo espaço gerado costuma ser denotado por V_j .

Logicamente, a partir das suposições em questão, pode-se concluir que, em se levando em conta a condição (2.17),

$$V_j \subset V_{j'}, \quad j' < j \quad (2.20).$$

Adicionalmente, se o sistema $\{\mathbf{f}_{j,k}\}_{j,k \in Z}$ gera $L^2(\mathfrak{R})$, tem-se

$$\bigcup_{j \in Z} V_j = L^2(\mathfrak{R}) \quad (2.21).$$

Ademais, lembrando (2.18),

$$\bigcap \{V_j, W_{j'}\} = \mathbf{f}, \quad j' \leq j \quad (2.22).$$

2.3.3. Representações de Funções em bases de Ondaletas

Uma vez que o sistema ortonormal $\{\mathbf{y}_{j,k}\}_{j,k \in Z}$ é base de $L^2(\mathfrak{R})$ (Daubechies (1992)), existem coeficientes $\{d_{j,k}\}_{j,k \in Z}$ de forma que uma possível representação de uma dada função $f(t)$ pertencente à este espaço, bem poderia ser

$$f(t) = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} d_{j,k} \mathbf{y}_{j,k}(t) \quad (2.23) \quad \text{onde} \quad d_{j,k} = \langle f(t), \mathbf{y}_{j,k}(t) \rangle = \int_{-\infty}^{+\infty} f(t) \mathbf{y}_{j,k}(t) dt.$$

A representação acima, apesar de ideal, não poderia ser implementada pelos meios disponíveis de computação dos coeficientes. Sendo assim, empiricamente, deveriam haver dois níveis de resolução $m, M, m \leq M$ e duas funções $n(j)$ e $N(j)$ tais que a mesma deveria ser substituída pela seguinte representação com número finito de coeficientes.

$$f(t) \approx \sum_{j=m}^M \sum_{k=n(j)}^{N(j)} d_{j,k} \mathbf{y}_{j,k}(t) \quad (2.24)$$

A representação (2.24) apresenta os inconvenientes de desconsiderar tanto os infinitos níveis de resolução menores que m quanto os também infinitos níveis de resolução maiores de M bem

como, para cada nível de resolução j efetivamente analisado, tanto os infinitos parâmetros de localização menores que $n(j)$ quanto os maiores que $N(j)$.

Muito embora a possibilidade de computação de coeficientes associados à bases para o espaço $L^2(\mathfrak{R})$ possa ser considerada como uma questão utópica, nem todos os inconvenientes referentes à representação em questão possuem caráter incontornável.

Em se levando em conta o sistema $\{\mathbf{f}_{j,k}\}_{j,k \in Z}$ ao invés do sistema $\{\mathbf{y}_{j,k}\}_{j,k \in Z}$, o inconveniente de se desconsiderar níveis de resolução maiores que M não mais existiria devido ao fato do sistema ortonormal $\{\mathbf{f}_{M,k}\}_{k \in Z}$ consistir conjunto gerador de $V_j, j > M$. Contudo, além da permanência das desconsiderações remanescentes, às mesmas se incorporaria, em primeira análise, o inconveniente de a representação em questão apresentar redundância no sentido em que proviria de um sistema linearmente dependente (*trade-off*).

O estabelecimento de um sistema $\{\mathbf{y}_{j,k}\}_{j,k \in Z}$ (isto é, de uma ondaleta-mãe $\mathbf{y}(t)$), não garante a existência de um sistema $\{\mathbf{f}_{j,k}\}_{j,k \in Z}$ correspondente. Todavia, se tal sistema existe, é possível obter-se representações para o espaço V_m que, além de encamparem a vantagem obtida através da substituição do Sistema I pelo Sistema II, se tornam, devido à ortonormalidade das mesmas, desprovidas de redundância. No caso, trata-se de representações ortonormais mistas (isto é, que envolvem elementos dos dois sistemas). Tal resultado é atingido devido às duas propriedades a seguir:

- Uma vez escolhida uma ondaleta-mãe para a qual existe o sistema normal $\{\mathbf{f}_{j,k}\}_{j,k \in Z}$, para qualquer nível de resolução j , existirá uma seqüência $\{h_k\}_{k \in Z}$ de forma que, sendo $h_k = (-1)^k l_{1-k}$, $\mathbf{y}_{j,k}(t) = \sum_{k \in Z} h_k \mathbf{f}_{j-1,k}$. Portanto, no caso, qualquer ondaleta referente a um nível de resolução j gerada a partir de $\mathbf{y}(t)$ pode ser escrita como combinação linear das ondaletas geradas a partir de $\mathbf{f}(t)$ referentes ao nível de resolução imediatamente inferior e, por conseqüência, a todos os níveis de resolução inferiores. Conseqüentemente,

$$W_j \subset V_{j'}, \quad j, j' \in Z, j' < j \quad (2.25).$$

- Uma vez definida uma ondaleta-mãe provida de sistema $\{\mathbf{f}_{j,k}\}_{j,k \in Z}$ correspondente pode-se demonstrar que, dado um nível de resolução genérico j , o conjunto formado através da unificação entre duas bases quaisquer para os espaços V_j e W_j constitui base para o espaço V_{j-1} , ou seja,

$$V_j = V_{j+1} \oplus W_{j+1} \quad (2.26).$$

Definidos os níveis de resolução “mínimo” (m) e “máximo” (M), a representação

$$f(t) \approx \sum_{j=m-1}^M \sum_{k=n(j)}^{N(j)} c_{j,k} \mathbf{f}_{j,k}(t) \quad (2.27)$$

pode ser escrita de maneira equivalente, tendo-se em vista os resultados anteriores, na forma

$$f(t) \approx \sum_{k=n(m)}^{N(m)} d_{m,k} \mathbf{y}_{m,k}(t) + \sum_{j=m}^M \sum_{k=n(j)}^{N(j)} c_{j,k} \mathbf{f}_{j,k}(t) \quad (2.28).$$

A partir de sucessivas iterações, obtém-se

$$f(t) \approx \sum_{j=m}^M \sum_{k=n(j)}^{N(j)} d_{j,k} \mathbf{y}_{j,k}(t) + \sum_{k=n(M)}^{N(M)} c_{M,k} \mathbf{f}_{M,k}(t) \quad (2.29)$$

onde

$$c_{M,k} = \langle f(t), \mathbf{f}_{M,k}(t) \rangle = \int_{-\infty}^{+\infty} f(t) \mathbf{f}_{M,k}(t) \quad (2.30) \quad , \quad d_{j,k} = \langle f(t), \mathbf{y}_{j,k}(t) \rangle = \int_{-\infty}^{+\infty} f(t) \mathbf{y}_{j,k}(t) \quad (2.31) .$$

Tal representação, com a qual se prosseguirá a análise, além de não desconsiderar, devido às propriedades acerca do sistema $\{\mathbf{f}_{j,k}\}_{k \in Z}$, níveis de resolução maiores que M , possui a característica de ser desprovida de redundância devido à propriedade (2.18).

2.3.4. Acerca de Alguns Pares $\{\mathbf{y}(t), \mathbf{f}(t)\}$ de suporte compacto

Muito embora haja ondaletas-mães para as quais não há ondaletas-pais associadas, um ponto importante consiste no fato de que, tendo-se em vista a satisfação das propriedades relativas às inter-relações entre os sistemas $\{\mathbf{y}_{j,k}\}_{k \in Z}$ e $\{\mathbf{f}_{j,k}\}_{k \in Z}$, diferentes ondaletas-mães deverão estar associadas, logicamente, à diferentes ondaletas-pais.

O primeiro par a ser considerado em várias publicações, devido à sua simplicidade relativa consiste em definir

$$\mathbf{y}(t) = \begin{cases} 1 & , 0 \leq t \leq \frac{1}{2} \\ -1 & , \frac{1}{2} < t \leq 1 \end{cases} \quad (2.32)$$

, comumente denominada *ondaleta de Haar*. A ondaleta-pai associada é dada por

$$\mathbf{f}(t) = 1 \quad , 0 \leq t \leq 1 \quad (2.33)$$

de forma que

$$\mathbf{y}_{j,k}(t) = \begin{cases} 2^{-\frac{j}{2}} & 2^j k \leq t \leq 2^j \left(k + \frac{1}{2} \right) \\ -2^{-\frac{j}{2}} & 2^j \left(k + \frac{1}{2} \right) < t \leq 2^j (k+1) \end{cases} \quad , \quad \mathbf{f}_{j,k}(t) = 2^{-\frac{j}{2}} \quad , 2^j k \leq t \leq 2^j (k+1) \quad (2.34) .$$

De maneira relativamente simples, verifica-se que

$$\begin{cases} h_0 = l_0 = l_1 = -h_1 = 1 \\ h_i = l_i = 0 \quad , i \notin \{0;1\} \end{cases} \quad (2.35) .$$

A ondaleta de Haar é um elemento da família de ondaletas-mães $\{\mathbf{y}^N\}_{N \in \mathbb{N}^+}$ denominadas *doublets* (assim denominadas em homenagem à sua formuladora Ingrid Daubechies). A mesma consiste na única ondaleta de tal família que possui expressão numérica explícita e caráter

simétrico. Cada elemento da família em questão, além de admitir ondaleta-pai correspondente, gera sistema de ondaletas $\{y_{j,k}^N\}_{j,k \in Z^+}$ ortogonal.

Outras famílias de ondaletas-mães $\{y^N\}_{N \in \mathbb{N}^+}$ introduzidas por Ingrid Daubechies bastante difundidas consistem nas *symlets* (assim denominadas por possuírem caráter próximo da simetria, ao contrário do que ocorre com as *daubelets*) e nas *coiflets* (assim denominadas por Ingrid Daubechies em homenagem à Coifman). Cada elemento de cada família em questão, além de possuir caráter próximo da simetria e admitir ondaleta-pai correspondente, gera sistema de ondaletas $\{y_{j,k}^{QN}\}_{\substack{j,k \in Z \\ Q \in \{S,C\}}}$ ortogonal.

2.3.5. A Transformada Discreta de Ondaletas

Além da desconsideração de determinados níveis de resolução e parâmetros de localização temporal, duas limitações técnicas adicionais merecem destaque:

- Devido à impossibilidade de se coletar dados acerca de determinado processo em intervalos infinitesimais de tempo, funções que apresentem caráter de continuidade, mesmo que apenas de forma local, não mais poderão consistir objetos de análise. Dever-se-á, portanto, considerar seqüências de quadrado somável de valores indexados pelo tempo, representadas de acordo com a forma $\{y_t\}_{t \in Z}$ e denominadas *séries de tempo*.
- Devido à impossibilidade de se coletar quantidade infinita de dados acerca de um determinado processo, seqüências de valores indexados por infinitos instantes de tempo também não poderão consistir objeto de análise. Dever-se-á, portanto, considerar séries de tempo cujo domínio de indexação esteja limitado à instantes mínimo (1) e máximo (T), representados de acordo com a forma $\{y_t\}_{t=1}^T$.

Logicamente, devido às necessidades de particularização do objeto de análise impostas pelas limitações em questão, o sistema de equações (2.30)-(2.31) se torna não mais pertinente por supor coleta de dados em instantes infinitesimais de tempo (ou seja, de forma contínua). Pode-se, em processo análogo ao relativo à análise de Fourier, considerar para o espaço das seqüências com T termos o produto interno dado por

$$\langle u, v \rangle = \sum_{t=1}^T u_t \cdot v_t \quad (2.36).$$

No caso, obtém-se a representação

$$c_{M,k} = \sum_{t=1}^T y_t f_{M,k}(t) \quad (2.37) \quad , \quad d_{j,k} = \sum_{t=1}^T y_t y_{j,k}(t) \quad (2.38)$$

denominada *transformada discreta de ondaletas de* $\{y_t\}_{t=1}^T$.

Um ponto importante a ser ressaltado surge acerca dos valores $N(j)$, $n(j)$, m e M . De acordo com o processo de dilatações binárias, partindo-se de uma função original de suporte compacto $w(t)$, observa-se que a medida dos suportes das funções pertencentes ao sistema $\{w_{j,k}\}_{k \in Z}$ (\wp_j) corresponde ao dobro do relativo ao nível de resolução anterior ($j-1$). Conseqüentemente,

$$\wp_j = 2^{(j-m)} \wp_0, \quad j \in Z \quad (2.39)$$

onde φ_0 corresponde à medida do suporte da função original. Sendo assim, tendo-se em vista relativa homogeneização dos domínios relativos aos sistemas $\{y_{j,k}\}_{k \in Z}$ e $\{f_{j,k}\}_{k \in Z}$, costuma-se considerar a representação

$$y_t \approx \sum_{j=m}^M \sum_{k=0}^{\left\lfloor \frac{T}{\varphi_j} \right\rfloor} d_{j,k} y_{j,k}(t) + \sum_{k=0}^{\left\lfloor \frac{T}{\varphi_M} \right\rfloor} c_{M,k} f_{M,k}(t) \quad (2.40).$$

Por outro lado, tendo-se em vista relativa homogeneidade entre a quantidade de dados e de coeficientes de ondaletas a serem computados, costuma-se considerar $m=1$ e $M=\lceil \log_2(T) \rceil$ ⁴.

Cabe notar que a aplicação do sistema de equações (2.37)-(2.38) não consiste em uma maneira computacionalmente ótima de obter tais coeficientes, em especial no caso em que T consiste em uma potência inteira de dois. A maioria dos programas computacionais os obtém através de algoritmos iterativos inspirados em um esquema de decimação binária proposto por Mallat (1989).

2.3.6. Funções Espectrais em Ondaletas

Dada uma série de tempo $\{y_t\}_{t=1}^T$, uma medida empírica de sua variabilidade (ou *energia*) bastante difundida em estatística consiste em sua soma de quadrados de desvios com relação à média, também denominada *soma de quadrados total*. A mesma é representada de acordo com a forma

$$Eng(\{y_t\}_{t=1}^T) = SQT(\{y_t\}_{t=1}^T) = \sum_{t=1}^T (y_t - \bar{y})^2 \quad (2.41)$$

onde $\bar{y} = \frac{\sum_{t=1}^T (y_t)}{T}$ corresponde à *média* da série.

Um ponto de extrema importância com relação à presente seção é que dada uma série de tempo $\{y_t\}_{t=1}^T$ razoavelmente aproximável pelo lado direito da expressão (2.29), sua energia total pode ser representada pela expressão

$$Eng(\{y_t\}_{t=1}^T) = \sum_{j=m}^M \sum_{k=n(j)}^{N(j)} d_{j,k}^2 \quad (2.42).$$

Adicionalmente, de acordo com (2.42), a mesma pode ser particionada tendo-se como base os diferentes pares (j,k) considerados. Definindo-se a *energia da série* $\{y_t\}_{t=1}^T$ com relação ao par $(j,k)_{j,k \in Z}$ de acordo com a forma

$$Eng_{j,k} = d_{j,k}^2 \quad (2.43)$$

⁴ Considera-se, no presente artigo, $\lfloor x \rfloor$ como sendo o maior inteiro menor ou igual à x e $\lceil x \rceil$ como sendo o menor inteiro maior ou igual à x .

obtém-se a expressão

$$Eng(\{y_t\}_{t=1}^T) = \sum_{j=m}^M \sum_{k=n(j)}^{N(j)} Eng_{j,k} \quad (2.44).$$

Analogamente ao que ocorre na análise de Fourier, a função $f(Z^2 \rightarrow \Re): (j, k) \rightarrow Eng_{j,k}$ é denominada *periodograma de ondaletas de* $\{y_t\}_{t=1}^T$.

A relação em questão permite que, alternativamente, séries $\{y_t\}_{t=1}^T$ tenham suas energias particionadas tendo-se como base os diferentes níveis de resolução j considerados pelo lado direito da expressão (2.29). Definindo-se a *energia da série* $\{y_t\}_{t=1}^T$ com relação ao nível de resolução $j \in Z$ de acordo com a forma

$$Eng_j = \sum_{k=n(j)}^{N(j)} d_{j,k}^2 \quad (2.45)$$

obtém-se a expressão

$$Eng(\{y_t\}_{t=1}^T) = \sum_{j=m}^M Eng_j \quad (2.46).$$

A função $f(Z \rightarrow \Re): j \rightarrow Eng_j$ é denominada *escalograma de ondaletas de* $\{y_t\}_{t=1}^T$.

3. Acerca dos Procedimentos

3.1. Redução de Ruído (*de-noising*)

O procedimento em questão aplicado à uma dada série de tempo $\{y_t\}_{t=1}^T$ atua na forma de uma eventual redução (sujeição à multiplicação por um número real pertencente ao intervalo [0;1]) de cada um dos elementos do sistema $\{d_{j,k}\}_{\substack{m \leq j \leq M \\ n(j) \leq k \leq N(j)}}$ de acordo com um determinado conjunto de critérios. Sua utilização leva a série a ser expressada na forma

$$y_t \approx \sum_{j=m}^M \sum_{k=n(j)}^{N(j)} \tilde{d}_{j,k} y_{j,k}(t) + \sum_{k=n(M)}^{N(M)} c_{M,k} f_{M,k}(t) \quad (3.1)$$

onde, $\forall (j, k) \left\{ \begin{array}{l} m \leq j \leq M \\ n(j) \leq k \leq N(j) \end{array} \right\}$, $\left\{ \begin{array}{l} d_{j,k} \cdot \tilde{d}_{j,k} \geq 0 \\ |\tilde{d}_{j,k}| \leq |d_{j,k}| \end{array} \right.$ e os coeficientes $\{\tilde{d}_{j,k}\}_{\substack{m \leq j \leq M \\ n(j) \leq k \leq N(j)}}$ são obtidos

através da aplicação de limiares aos coeficientes originais $\{d_{j,k}\}_{\substack{m \leq j \leq M \\ n(j) \leq k \leq N(j)}}$.

Com relação aos limiares, as questões referentes aos seus procedimentos de escolha e ao algoritmo de redução dos coeficientes $\{d_{j,k}\}_{\substack{m \leq j \leq M \\ n(j) \leq k \leq N(j)}}$ assumem fundamental relevância. Com relação ao segundo paradigma, resultam bastante difundidos dois algoritmos de redução (denominados, respectivamente, de *duro* (*hard threshold*) e *suave* (*soft threshold*)) nos meios estatísticos. Uma vez definido um dado limiar I , os algoritmos duro e suave podem ser expressos de maneira respectiva de acordo com as formas (3.2) e (3.3)

$$\tilde{d}_{j,k}^D = \begin{cases} 0, & |d_{j,k}| \leq I \\ d_{j,k}, & |d_{j,k}| > I \end{cases} \quad (3.2) \quad \tilde{d}_{j,k}^S = \begin{cases} 0, & |d_{j,k}| \leq I \\ s(d_{j,k}) \cdot (|d_{j,k}| - I), & |d_{j,k}| > I \end{cases} \quad (3.3)$$

, onde s representa a função sinal. A escolha entre os mesmos está relacionada a um *trade-off* entre viés (suave) e variância (duro) associado ao processo de redução de ruído em questão.

Com relação ao primeiro paradigma, resultam bastante difundidos dois procedimentos de escolha denominados, respectivamente, de *universal* e *adaptativo*. O procedimento universal consiste no estabelecimento de um único limiar I , relacionado ao desvio-padrão \mathbf{s} do sistema $\{d_{j,k}\}_{\substack{m \leq j \leq M \\ n(j) \leq k \leq N(j)}}$, para todos os elementos pertencentes ao mesmo. O procedimento adaptativo consiste na discriminação entre limiares de acordo com os níveis de resolução presentes no sistema em questão de forma a estabelecer, para cada nível de resolução $m \leq j \leq M$, um limiar I_j relacionado ao desvio-padrão \mathbf{s}_j do sistema $\{d_{j,k}\}_{n(j) \leq k \leq N(j)}$.

No contexto em questão, um ponto importante a ser ressaltado consiste no método de estimação dos desvios-padrão associados aos coeficientes $\{d_{j,k}\}_{\substack{m \leq j \leq M \\ n(j) \leq k \leq N(j)}}$. Ao se considerar o procedimento universal de escolha resultam bastante difundidos dois métodos (que procedem a estimação de \mathbf{s} tendo-se como base, respectivamente, o sistema $\{d_{j,k}\}_{\substack{m \leq j \leq M \\ n(j) \leq k \leq N(j)}}$ e o sistema $\{d_{m,k}\}_{n(m) \leq k \leq N(m)}$). Já ao se considerar o procedimento adaptativo a estimativa do desvio-padrão de um dado coeficiente $d_{j,k,m \leq j \leq M}$ será definida pelo nível de resolução associado ao mesmo.

As expressões de limiares universais e adaptativos mais difundidas podem ser encontradas em Morettin (1997).

3.2. O Fracionamento da Série de Tempo

3.2.1. A Problemática

Formalmente, a problemática corresponde à obtenção de uma frequência crítica I_y , de maneira que, enquanto componentes de $\{y_t\}_{t=1}^T$ relacionados à frequências menores ou iguais a I_y devam formar a sub-série $\{x_t\}_{t=1}^T$ relativa às baixas frequências, os demais componentes de $\{y_t\}_{t=1}^T$ formem a sub-série relativa às altas frequências, denotada $\{z_t\}_{t=1}^T$.

Em relação às séries econômicas pode-se obter uma forma de se escolher I_y , descrita a seguir, baseada na observação de que o comportamento cíclico das mesmas costuma ser formado por dois movimentos. Enquanto um deles seria de curto-prazo (o qual se repete em intervalos curtos de tempo, podendo, portanto, ser caracterizado como sendo de baixo período e, conseqüentemente, de alta frequência), o outro seria de longo-prazo, que pode ser classificado como sendo de alto período e, portanto, de baixa frequência.

3.2.2. O Procedimento

No caso em que o fato supracitado se verifica para uma determinada série de tempo $\{y_t\}_{t=1}^T$ é razoável imaginar que as frequências relativas aos dois movimentos devam concentrar maiores porções da energia total da série que as demais. Levando-se em conta a existência de relação direta entre frequência e nível de resolução em análise de ondaletas, se torna razoável esperar que o

escalograma de $\{y_t\}_{t=1}^T$ apresente seus picos em dois níveis de resolução p e P , $m \leq p \leq P \leq M$. No caso, p se relaciona ao movimento de longo-prazo e P ao movimento de curto-prazo da série.

Considerando p e P como sendo índices “representativos” dos níveis de resolução associados, respectivamente, aos componentes de baixas e altas frequências da série, Ariño e Vidakovic (1995) sugerem que, no caso em que os mesmos apresentem diferentes caracteres de paridade, proceda-se o fracionamento da série

$$y_t = \sum_{j=m}^M \sum_{k=n(j)}^{N(j)} d_{j,k} y_{j,k}(t) + \sum_{k=n(M)}^{N(M)} c_{M,k} f_{M,k}(t) \quad (3.4)$$

de acordo com a forma

$$x_t = \sum_{k=n(M)}^{N(M)} c_{M,k} f_{M,k}(t) + \sum_{j=[a]}^M \sum_{k=n(j)}^{N(j)} d_{j,k} y_{j,k}(t) \quad (3.5) \quad , \quad z_t = \sum_{j=m}^{[a]} \sum_{k=n(j)}^{N(j)} d_{j,k} y_{j,k}(t) \quad (3.6)$$

onde $a = \frac{p+P}{2} \in Z$. Sugere-se, portanto, que o critério de classificação de um determinado nível de resolução j como sendo associado à baixa ou alta frequências seja baseado em sua *distância numérica* com relação aos considerados “representativos” de tais características em $\{y_t\}_{t=1}^T$.

No caso em questão observa-se que, a partir da adoção do mesmo, as sub-séries $\{x_t\}_{t=1}^T$ e $\{z_t\}_{t=1}^T$ apresentam duas propriedades:

- Sua soma corresponde exatamente à série original $\{y_t\}_{t=1}^T$;
- A soma entre suas energias $Eng(\{x_t\}_{t=1}^T)$ e $Eng(\{z_t\}_{t=1}^T)$ corresponde exatamente à energia $Eng(\{y_t\}_{t=1}^T)$ relativa à série original.

No caso em que p e P possuíssem o mesmo caráter de paridade, devido à duplicidade de representação do nível de resolução $a = \frac{p+P}{2} \in Z$, à menos que todos os coeficientes de ondaletas referentes ao mesmo fossem nulos, nenhuma das propriedades supracitadas se manteria válida.

Muito embora tais propriedades possam ser recuperadas adaptando-se o procedimento em questão de forma que o mesmo, além de discriminar entre diferentes níveis de resolução, faça o mesmo para diferentes valores de k para o nível de resolução a , tal adaptação incorreria no absurdo de diferenciar frequências relativas à ondaletas que as possuem idênticas.

Apesar de, salvo casos especiais, não ser possível recuperar de forma conjunta as propriedades em questão através de uma adaptação que, ao invés de discriminar diferentes valores de k , destine, para cada uma das sub-séries, frações de cada um dos coeficientes relativos ao nível de resolução a , de forma que

$$\begin{cases} x_t = \sum_{k=n(a)}^{N(a)} a_k \cdot d_{a,k} y_{a,k}(t) + \sum_{j=a+1}^M \sum_{k=n(j)}^{N(j)} d_{j,k} y_{j,k}(t) + \sum_{k=n(M)}^{N(M)} c_{M,k} f_{M,k}(t) & (3.9) \\ z_t = \sum_{j=m}^{a-1} \sum_{k=n(j)}^{N(j)} d_{j,k} y_{j,k}(t) + \sum_{k=n(a)}^{N(a)} b_k \cdot d_{a,k} y_{a,k}(t) & (3.10) \end{cases}$$

, a mesma possuiria a vantagem de evitar o problema acerca da adaptação anterior. Nesse sentido, Ariño e Vidakovic (1995) sugerem duas formas alternativas, descritas a seguir, de se proceder ao fracionamento dos coeficientes em questão, cada uma possuindo uma de tais propriedades.

Fracionamento Preservador dos Valores Numéricos da Série Original

Formalmente, devido ao caráter de ortonormalidade do sistema $\{y_{j,k}\}_{j,k \in \mathbb{Z}}$, um fracionamento $\{a_k, b_k\}_{k=n(a)}^{N(a)}$ preservará os valores numéricos da série original se e somente se, $\forall k \in \{n(a), \dots, N(a)\}$, $a_k + b_k = 1$.

Pode-se, de forma aproximada, afirmar que o domínio temporal relativo à uma dada ondaleta $y_{j,k}$ equivalha tanto ao da ondaleta $y_{j+1, \lfloor \frac{k}{2} \rfloor}$ quanto ao do sistema $\{y_{j-1,2k}, y_{j-1,2k+1}\}$.

Dessa forma, uma maneira de se fracionar um dado coeficiente $d_{a,k}$ poderia consistir em se considerar a_k e b_k de forma a refletir, respectivamente, as contribuições relativas aos “fluxos” de “entrada” e “saída” de energia associadas ao suporte de $y_{a,k}$ tendo-se como direcionamento o sentido crescente do grau de resolução para a energia associada ao par (a, k) , ou seja,

$$a_k = \frac{a}{a+b} \quad (3.11) \quad , \quad b_k = 1 - a_k = \frac{b}{a+b} \quad (3.12)$$

, onde

$$a = d_{j+1, \lfloor \frac{k}{2} \rfloor} \quad (3.13) \quad , \quad b = \frac{d_{j-1,2k}^2 + d_{j-1,2k+1}^2}{2} \quad (3.14)$$

Fracionamento Preservador da Energia da Série Original

Formalmente, qualquer fracionamento $\{a_k, b_k\}_{k=n(a)}^{N(a)}$ tal que, $\forall k \in \{n(a), \dots, N(a)\}$,

$$a_k^2 + b_k^2 = 1 \quad (3.15)$$

preservará a energia relativa à série original.

Levando-se em conta as considerações relativas à seção anterior bem como a expressão (3.15), raciocínio análogo pode ser desenvolvido de forma a se obter expressões similares para o caso em questão. Ao se considerar a_k e b_k de forma que os mesmos reflitam as raízes das contribuições relativas ao caso anterior (logicamente positivas, por representarem grandezas quadradas) associadas ao suporte de $y_{a,k}$ tendo-se como direcionamento o sentido crescente do grau de resolução para a energia associada ao par (a, k) , obtém-se

$$a_k = \sqrt{\frac{a}{a+b}} \quad (3.16) \quad , \quad b_k = \sqrt{1 - a_k^2} = \sqrt{\frac{b}{a+b}} \quad (3.17)$$

$$\text{, onde} \quad 0 \leq a = d_{j+1, \lfloor \frac{k}{2} \rfloor} \quad (3.18) \quad , \quad 0 \leq b = \frac{d_{j-1,2k}^2 + d_{j-1,2k+1}^2}{2} \quad (3.19).$$

4. Aplicação dos Procedimentos

Para a aplicação dos procedimentos relacionados à modelagem das séries, optou-se por se utilizar os dados referentes aos cento e vinte e oito primeiras observações. Escolheu-se ainda a ondaleta de Haar e o procedimento universal de escolha de acordo com a expressão

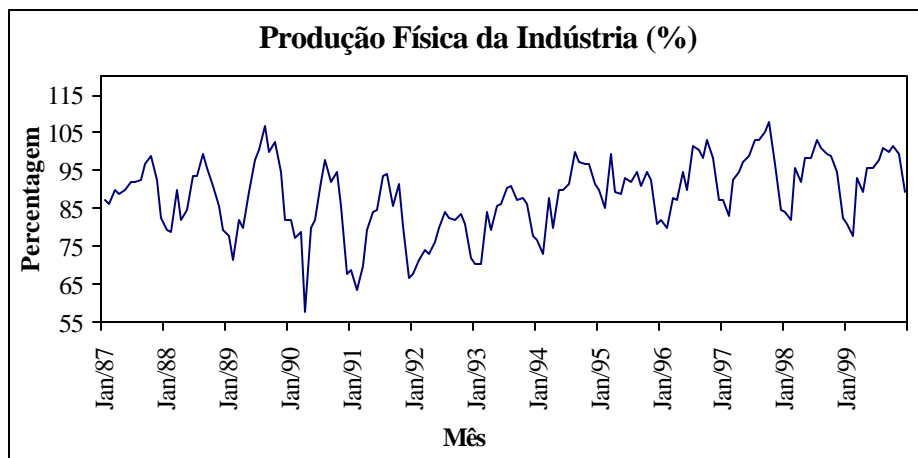
$$I = s \sqrt{2 \ln(T)} \quad (4.1)$$

onde s é estimado tendo-se como base a integralidade do sistema $\{d_{j,k}\}_{\substack{1 \leq j \leq 7 \\ 0 \leq k \leq -1 + 2^{j-1}}}$. Utilizou-se o algoritmo duro de redução. A obtenção dos coeficientes associados às ondaletas de Haar (cujas magnitudes estão representadas por intensidade de luminosidade nos gráficos anexos A.1 à A.6) e o processo de retirada de ruído foram implementados através do programa *MATLAB* v. 5.3 r.11. Já o processo de modelagem Box-Jenkins foi implementado através do programa *Eviews* v. 3.0.

4.1. Série da Produção industrial (*ProdInd*)

Esta série é composta por cento e cinquenta e seis dados mensais referentes à produção física industrial brasileira com base fixa em agosto de 1994. O Gráfico 1 apresenta a série *ProdInd* para o período de janeiro de 1987 a dezembro de 1999.

Gráfico 1



Fonte: IBGE

Através da aplicação da metodologia Box-Jenkins (procedimento tradicional), ajustou-se o modelo ARIMA $(0,1,1)(2,0,0)_6$ à série em questão. Os resultados são apresentados na equação (4.2).

$$(1 - L) \left(1 + \underset{(0,080402)}{0.260285 L^6} - \underset{(0,080873)}{0.529737 L^{12}} \right) \text{ProdInd}_t = (1 + \underset{(0,089302)}{0.366585 L}) e_t \quad (4.2)$$

onde $e_t \sim IIN(0, 25.68293)$.

Conforme apresentado na seção 3.2.2, a implementação do procedimento AV, através da análise do escalograma, apresentado na Tabela 1, à série *ProdInd*, gera o seguinte fracionamento em baixas (4.3) e altas (4.4) frequências

$$x_t = c_{7,0} \cdot f_{7,0}(t) + \sum_{j=3}^7 \sum_{k=0}^{-1+2^{7-j}} d_{j,k} \cdot y_{j,k}(t) \quad (4.3) \quad , \quad z_t = \sum_{j=1}^2 \sum_{k=0}^{-1+2^{7-j}} d_{j,k} \cdot y_{j,k}(t) \quad (4.4)$$

Este fracionamento encontra-se representado nos Gráficos 2 e 3.

Tabela 1

j	1	2	3	4	5	6	7
Eng_j	1355.1	2384.2	3598.6	809.0	1187.1	1844.5	325.1

Gráfico 2

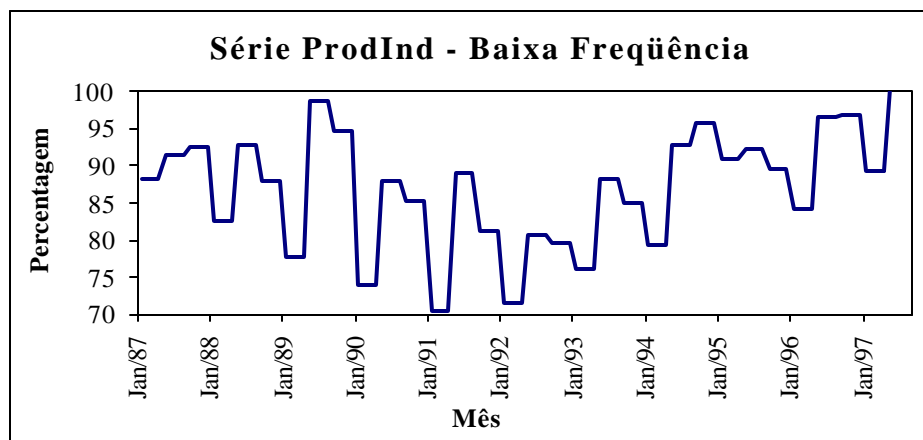
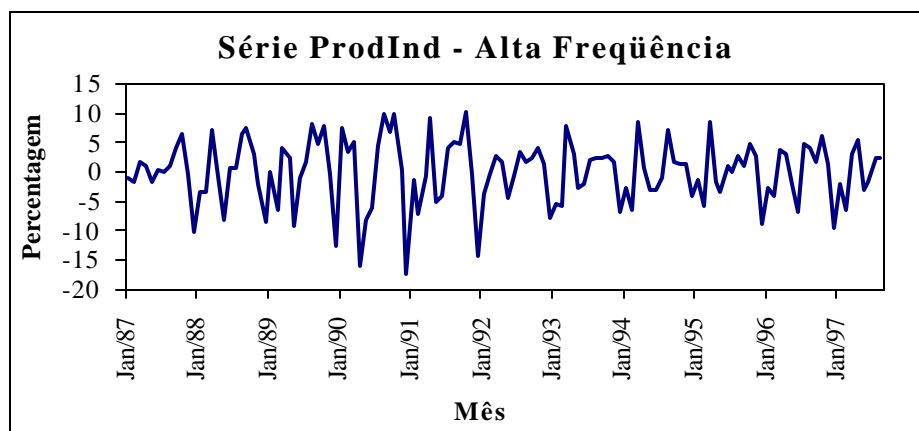


Gráfico 3



Aplicando-se a metodologia Box-Jenkins às séries fracionadas $\{x_t\}_{t=1}^{128}$ e $\{z_t\}_{t=1}^{128}$, ajustou-se os modelos ARIMA (4.5) e (4.6).

$$(1-L) \left(1 + \frac{0.123061L^4}{(0.044397)} - \frac{0.841319L^{12}}{(0.045914)} \right) x_t = \left(1 + \frac{0.484996L^8}{(0.066430)} + \frac{0.420050L^{12}}{(0.066444)} \right) h_t \quad (4.5)$$

$$(1-L^{12}) \left(1 + \frac{0.501327L^{12}}{(0.091452)} \right) z_t = \left(1 + \frac{0.285165L}{(0.098899)} + \frac{0.477922L^2}{(0.100124)} + \frac{0.181385L^3}{(0.105707)} + \frac{0.300751L^4}{(0.101432)} - \frac{0.222985L^{11}}{(0.066474)} \right) x_t \quad (4.6)$$

onde $h_t \sim \text{IIN}(0,5.304764)$ (4.7) e $x_t \sim \text{IIN}(0,11.17904)$ (4.8)

Cabe ressaltar que a caracterização dos modelos ARIMA ajustados para cada uma das sub-séries consiste em indício de que o procedimento de fracionamento aplicado obteve relativo êxito em discriminar as frequências da série. De acordo com o mesmo, enquanto a sub-série $\{x_t\}_{t=1}^{128}$ representa a evolução de longo prazo da série original a sub-série $\{z_t\}_{t=1}^{128}$ apresenta, embora de forma não pura, seu comportamento sazonal.

Através da aplicação da metodologia Box-Jenkins à série alisada, isto é, depois de realizado o *de-noising* da série original, ajustou-se o modelo ARIMA(0,1,2)(2,0,0) com sazonalidade semestral.

$$(1-L)\left(1 + \frac{0.271005}{(0.084273)}L^6 - \frac{0.458812}{(0.08458)}L^{12}\right) \text{ProdInd}_t = \left(1 + \frac{0.283811}{(0.090879)}L^2\right) e_t \quad (4.9)$$

onde $e_t \sim IIN(0, 19.6564)$.

A série alisada é mostrada no Gráfico 4, enquanto o Gráfico 5 apresenta os valores realizados e previstos fora da amostra para os três procedimentos utilizados (tradicional, AV e *de-noising*). Os erros absolutos médios percentuais da série original (*MAPE's*) são apresentados na Tabela 5.

Gráfico 4

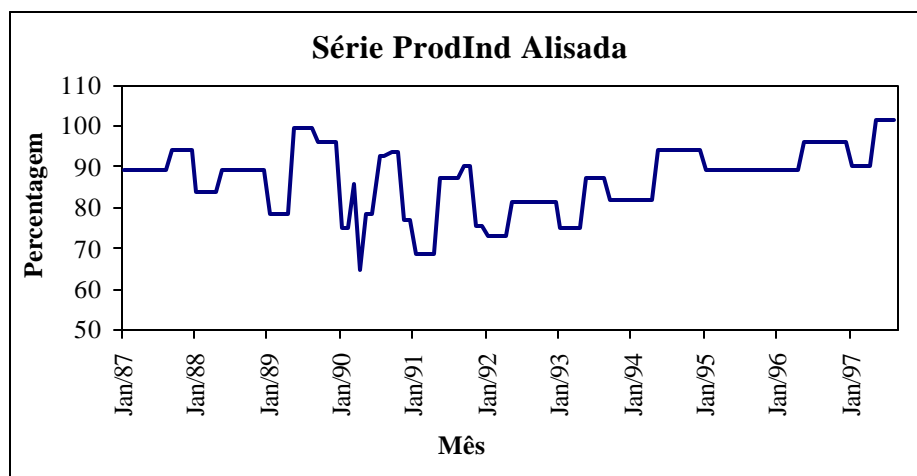
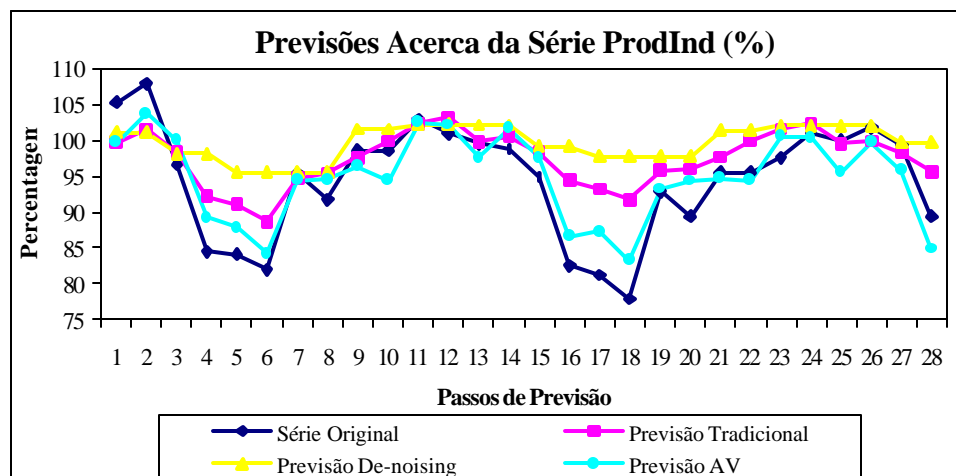


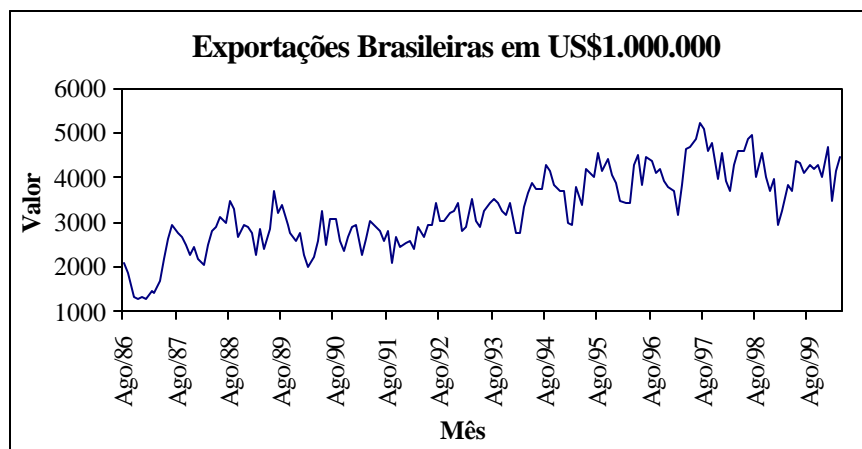
Gráfico 5



4.2. Série de Exportações Brasileiras (*Expor*)

Esta série é composta por cento e sessenta e quatro dados mensais referentes às exportações brasileiras em milhões de dólares para o período que vai de agosto de 1986 a março de 2000, sendo representada pelo Gráfico 6.

Gráfico 6



Fonte: Ipeadata.

Através da aplicação da metodologia Box-Jenkins (tradicional), ajustou-se o modelo ARIMA(0,1,6)(0,0,1) com sazonalidade mensal.

$$(1 - L)(1 - \underset{(0.081800)}{0.268272} L^{11} - \underset{(0.083475)}{0.368489} L^{12})Expor_t = (1 + \underset{(0.072040)}{0.622849} L + \underset{(0.071975)}{0.191806} L^6)e_t \quad (4.10)$$

onde $e_t \sim IIN(0, 86729.36)$.

A Tabela 2 mostra o escalograma relativo à *Expor* que é utilizado para obter-se o nível de corte (fracionamento) entre a alta e baixa frequência da série. Este fracionamento resultou nas equações (4.11) e (4.12), utilizadas a seguir para a implementação do procedimento AV. As sub-séries de baixas e alta frequência são apresentadas nos Gráficos 7 e 8.

$$x_t = c_{7,0} \cdot f_{7,0}(t) + d_{7,0} \cdot y_{7,0}(t) \quad (4.11) \quad , \quad z_t = \sum_{j=1}^6 \sum_{k=0}^{-1+2^{7-j}} d_{j,k} \cdot y_{j,k}(t) \quad (4.12)$$

Tabela 2

<i>j</i>	1	2	3	4	5	6	7
<i>Eng_j</i>	4838630.5	4253933.7	6270643.1	5884977.4	5938172.0	10929006.6	30543951.0

Gráfico 7

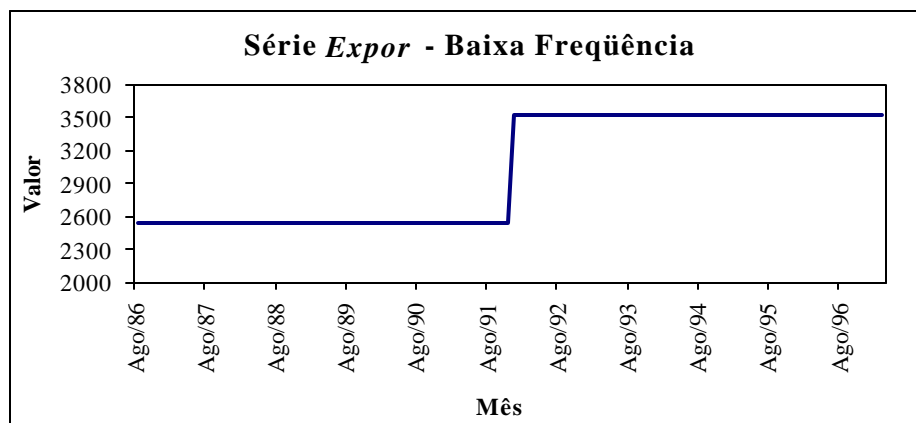
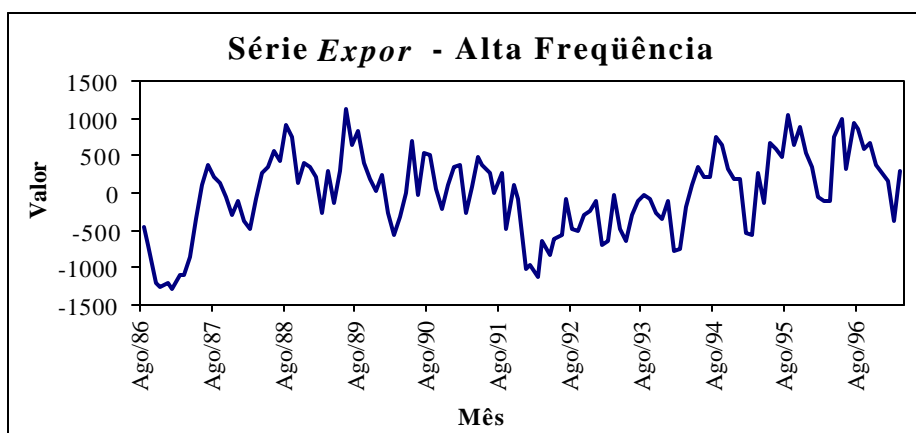


Gráfico 8



Através da aplicação da metodologia Box-Jenkins às séries $\{x_t\}_{t=1}^{128}$ e $\{z_t\}_{t=1}^{128}$, ajustou-se os modelos ARIMA.

$$(1-L)x_t = h_t \quad (4.13)$$

$$\left\{ \left(1 - \frac{0.561115}{(0.07095)} L - \frac{0.257375}{(0.06414)} L^{11} \right) z_t = \left(1 - \frac{0.232405}{(0.075387)} L^3 - \frac{0.202597}{(0.078763)} L^5 + \frac{0.402046}{(0.000799)} L^{30} \right) x_t \right. \quad (4.14)$$

$$\text{onde } h_t \sim \text{IIN}(0,7515.735) \quad (4.15), \quad x_t \sim \text{IIN}(0,83533.88) \quad (4.16).$$

Neste caso, ao contrário do que ocorreu para a série de produção industrial, os resultados dos modelos ARIMA ajustados para cada uma das sub-séries mostra que o fracionamento foi capaz de discriminar perfeitamente a parte estacionária da parte não estacionária da série. Enquanto a sub-série $\{x_t\}_{t=1}^{128}$ representa a evolução de longo prazo da série original (parte não estacionária), a sub-série $\{z_t\}_{t=1}^{128}$ apresenta seu comportamento sazonal (parte estacionária). Note-se que este procedimento pode ser utilizado também como um índice de que houve uma mudança estrutural na série. Como o escalograma da série mostra uma grande variação de energia nos níveis de resolução mais altos (6 e 7), tem-se um índice de que a série tem uma quebra estrutural que a divide em duas sub-amostras de igual tamanho. Se o objetivo fosse procurar por quebras estruturais, poder-se-ia continuar a calcular novos escalogramas para as sub-amostras até que os valores das energias se tornassem próximos.

Através da aplicação da metodologia Box-Jenkins à série alisada (*de-noised*), apresentada no Gráfico 9, ajustou-se um modelo ARIMA(0,1,0)(1,0,1) com sazonalidade mensal.

$$(1 - L) \left(1 + \underset{(0.068581)}{0.489254 L^4} - \underset{(0.064332)}{0.469292 L^{12}} \right) \text{Expôr} = \left(1 + \underset{(0.000433)}{0.429475 L^8} - \underset{(0.049628)}{0.473335 L^{12}} \right) e_t \quad (4.17)$$

onde $e_t \sim IIN(0, 7623.162)$. O confronto entre a série original e as previsões associadas aos três métodos encontram-se no Gráfico 10, enquanto que os erros absolutos médios em percentagem são apresentados na Tabela 5.

Gráfico 9

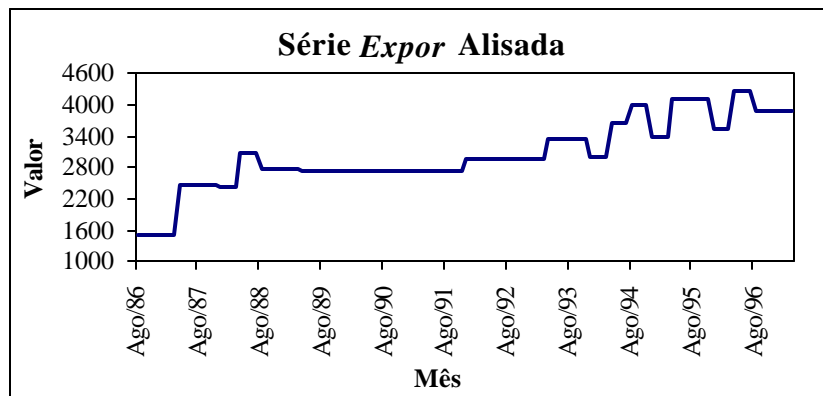
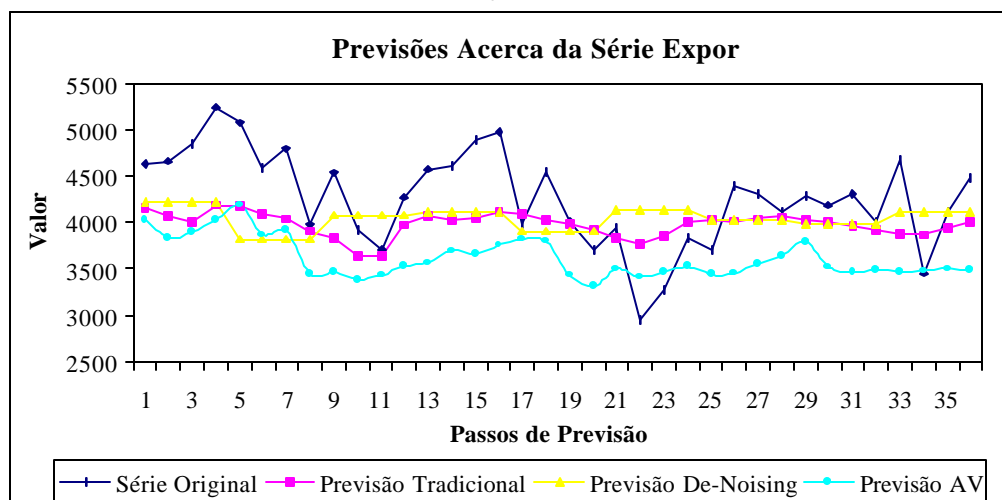


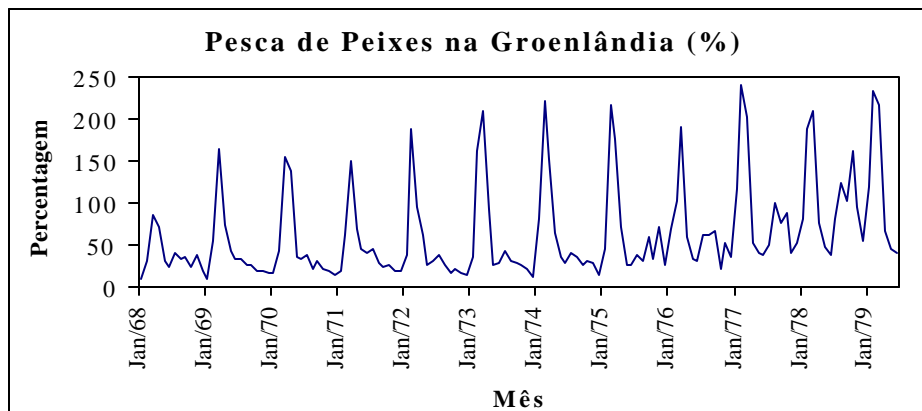
Gráfico 10



4.3. Série Volume de Pesca na Groelândia (*Fish*)

Esta série é composta por cento e trinta e oito dados mensais referentes à pesca de peixes na Groenlândia, com base fixa na média do ano de 1990. O Gráfico 11 apresenta a série *Fish* para o período de janeiro de 1968 a junho de 1979.

Gráfico 11



Fonte: FMI.

Através da aplicação da metodologia Box-Jenkins (tradicional), ajustou-se o seguinte modelo ARIMA(0,1,0)(1,1,1) com sazonalidade mensal.

$$(1-L)(1-L^{12}) \left(1 + \frac{0.184306L^3}{(0.09183)} + \frac{0.335063L^{12}}{(0.09432)} \right) \ln(Fish_t) = \frac{0.001864}{(0.000359)} (1 + \frac{1.059678L}{(0.020986)} - \frac{0.063941L^{12}}{(0.020878)}) e_t \quad (4.18)$$

onde $e_t \sim IIN(0, 0.110064)$.

A Tabela 3 apresenta o escalograma relativo à série *Fish*. Optou-se pela utilização do fracionamento preservador dos valores numéricos da série original e não de sua energia. As séries fracionadas de baixas e altas frequências são apresentadas nas equações (4.19) e (4.20), respectivamente. Os gráficos 12 e 13 apresentam estas duas séries.

$$\begin{cases} x_t = c_{7,0} \cdot f_{7,0}(t) + \sum_{j=3}^6 \sum_{k=0}^{-1+2^{7-j}} d_{j,k} \cdot y_{j,k}(t) + \sum_{k=0}^{31} a_k \cdot d_{2,k} \cdot y_{2,k}(t) & (4.19) \\ z_t = \sum_{k=0}^{31} (1-a_k) \cdot d_{2,k} \cdot y_{2,k}(t) + \sum_{k=0}^{63} d_{1,k} \cdot y_{1,k}(t) & (4.20) \end{cases}$$

A Tabela 4 apresenta os coeficientes $\{a_k\}_{k=0}^{31}$ relacionados ao segundo nível de resolução que foram sub-divididos entre os dois níveis de frequências.

Tabela 3

j	1	2	3	4	5	6	7
Eng_j	118169.3	46779.1	128464.5	33617.4	6265.7	14451.1	14822.1

Tabela 4

i	1	2	3	4	5	6	7	8	9
a_i	0.782	0.963	0.973	0.632	0.871	0.970	0.965	0.990	0.996
10	11	12	13	14	15	16	17	18	19
0.731	0.897	0.998	0.594	0.996	1.000	0.756	0.872	0.940	0.726
20	21	22	23	24	25	26	27	28	29
0.999	0.997	0.677	0.986	0.448	0.601	0.999	0.976	0.721	0.160
30	31	32							
0.641	0.536	0.944							

Gráfico 12

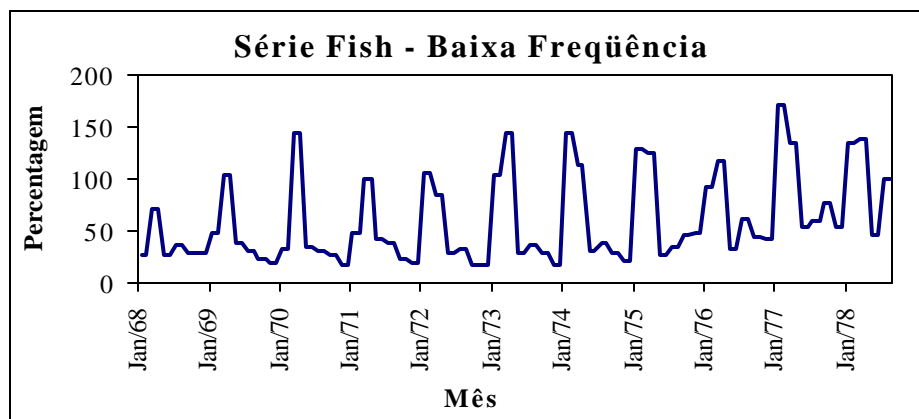
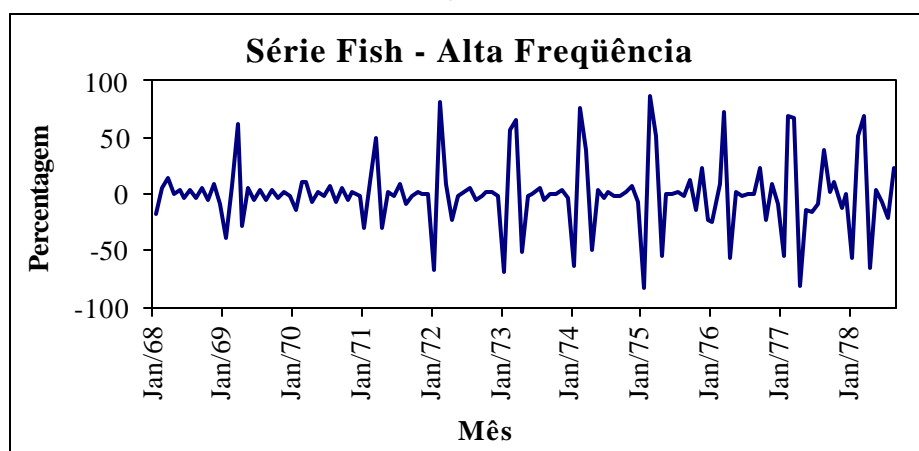


Gráfico 13



Através da aplicação da metodologia Box-Jenkins às séries fracionadas $\{x_t\}_{t=1}^T$ e $\{z_t\}_{t=1}^T$, ajustou-se os modelos ARIMA

$$\left\{ \begin{array}{l} (1-L)(1-L^{12}) \left(1 + \frac{0.439107 L^{12}}{(0.094528)} \right) x_t = \left(1 + \frac{0.959835 L^2}{(0.015079)} \right) h_t \end{array} \right. \quad (4.21)$$

$$\left\{ \begin{array}{l} (1-L^{12}) \left(1 + \frac{0.571962 L^{12}}{(0.083718)} \right) z_t = \left(1 + \frac{0.940479 L}{(0.075167)} + \frac{0.078728 L^3}{(0.078475)} \right) x_t \end{array} \right. \quad (4.22)$$

onde $h_t \sim \text{IIN}(0,227.3489)$ (4.23) e $x_t \sim \text{IIN}(0,133.6154)$ (4.24).

Observa-se que a série de baixas frequências $\{x_t\}_{t=1}^{128}$ acabou captando as diferenças tanto sazonal como consecutiva, enquanto que a série de altas frequências $\{z_t\}_{t=1}^{128}$ incorporou o padrão sazonal da série.

A aplicação da metodologia Box-Jenkins à série alisada (*de-noised*), representada no Gráfico 14, levou à obtenção de um modelo ARIMA(0,1,1)(1,1,0), apresentado na equação (4.25)

$$(1-L)(1-L^{12}) \left(1 + \frac{0.519918 L^{12}}{(0.085234)} \right) \text{Fish}_t = \left(1 + \frac{0.967718 L}{(0.023747)} \right) e_t \quad (4.25)$$

, onde $e_t \sim IIN(0,808.8815)$. O Gráfico 15 apresenta os valores para a série original e para as previsões pelos três métodos e os erros absolutos médios em percentagem estão na Tabela 5.

Gráfico 14

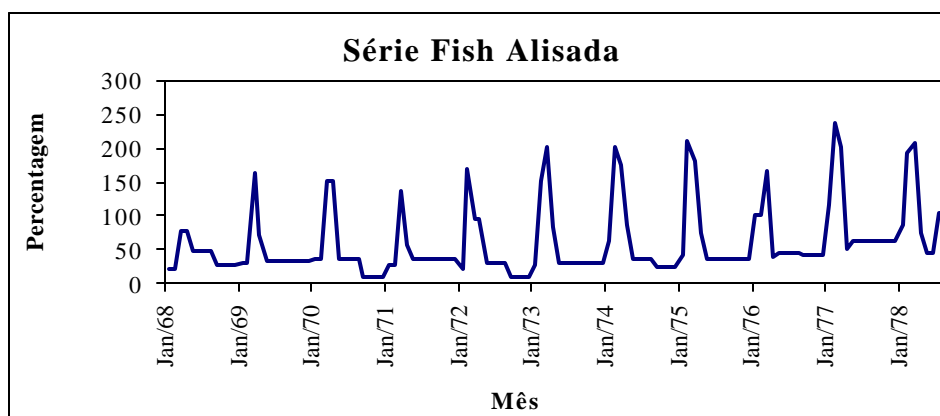


Gráfico 15

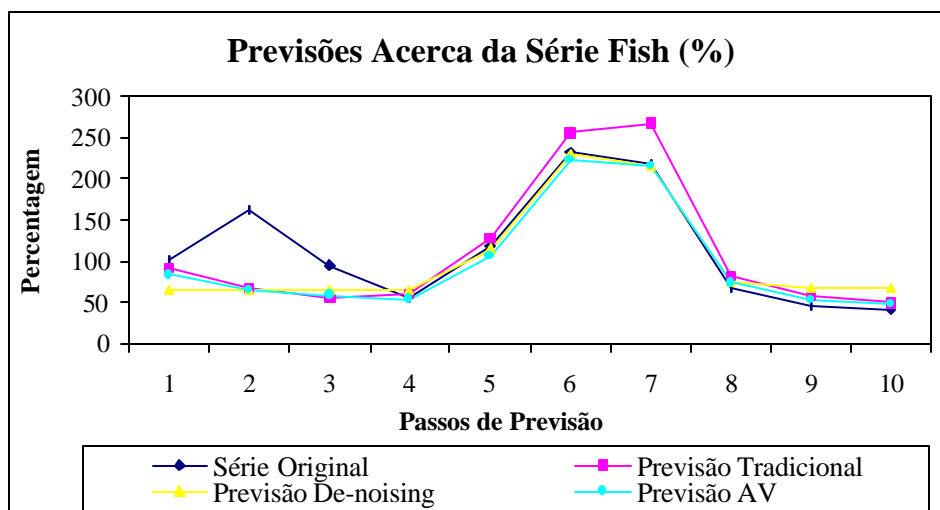


Tabela 5

Série	Procedimento	MAPE (%)
<i>ProdInd</i>	Tradicional	4.77909
<i>ProdInd</i>	AV	3.24786
<i>ProdInd</i>	De-noising	6.88524
<i>Expor</i>	Tradicional	9.95489
<i>Expor</i>	AV	15.52875
<i>Expor</i>	De-noising	10.72856
<i>Fish</i>	Tradicional	22.62733
<i>Fish</i>	AV	17.47298
<i>Fish</i>	De-noising	27.35032

5. Conclusões

Este artigo apresentou de forma sistemática uma introdução às ondaletas e dois possíveis métodos alternativos de previsão. Estes métodos consistem basicamente na aplicação da metodologia ARIMA a sub-séries extraídas da série original através das ondaletas. São utilizados o procedimento de alisamento (*de-noising*) e de divisão em séries de alta e baixa frequência.

Tendo-se como base os *MAPE's* referentes às três alternativas de procedimentos de modelagem relativamente às séries de tempo analisadas observa-se que o procedimento AV mostrou-se superior aos demais para duas das mesmas. Isto é um indício de que a modelagem em separado de sub-séries de baixas e altas frequências contribui positivamente para a qualidade de previsão de séries de tempo. O fato de o procedimento de alisamento ter se mostrado sempre inferior ao método tradicional constitui indício de que este primeiro, além de não contribuir positivamente para a qualidade de previsão de séries de tempo, possa estar associado à efeito em sentido contrário.

Cabe ressaltar que a análise conduzida no presente artigo carece de poder estatístico tanto no sentido em que um conjunto de três séries de tempo consiste em uma amostra muito pequena para uma eventual análise de proporções de relacionamentos entre variáveis aleatórias *MAPE* quanto no sentido em que a mesma não as considera de forma inferencial.

Referências

- Ariño, M. A. (1995), *Time Series Forecasts via Wavelets: An Application to Car Sales in The Spanish Market*, Discussion Paper 95-30, ISDS, Duke University.
- Ariño, M. A., Vidakovic, B. (1995), *On Wavelet Scalograms and Their Application in Economic Time Series*, Discussion Paper 95-21, ISDS, Duke University.
- Daubechies, I. (1988), *Orthonormal Bases of Compactly Supported Wavelets*, Comm. Pure Appl. Math., 41, 909-996.
- Daubechies, I. (1992), *Ten Lectures on Wavelets*, Philadelphia: SIAM.
- Donoho, D.L., Johnstone, I.M. (1992a), *Ideal Spatial Adaptation via Wavelet Shrinkage*, Technical Report, Stanford University.
- Donoho, D.L., Johnstone, I.M. (1992b), *Minimax Estimation via Wavelet Shrinkage*, Technical Report, Stanford University.
- Donoho, D.L., Johnstone, I.M. (1993), *Adapting to Unknown Smoothness via Wavelet Shrinkage*, Technical Report, Stanford University.
- Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., Picard, D. (1993), *Wavelet Shrinkage: Asymptopia?*, Technical Report, Stanford University.
- Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., Picard, D. (1995), *Universal near Minimaxity of Wavelet Shrinkage*, Technical Report, Stanford University.
- Johnstone, I.M., Silverman, B. (1996), *Wavelet Threshold Estimators for data with Correlated Noise*, Technical Report, Stanford University.
- Mallat, S. (1989), *A Theory for Multiresolution Signal Decomposition: Wavelet Representation*, IEEE Trans. Pattern Anal. Mach. Intellig., 11, 674-693.
- Meyer, Y. (1993), *Wavelets: Algorithms and Applications*. Philadelphia: SIAM.
- Morettin, P. (1997), *Ondaletas e seus usos na Estatística*. 7^a Escola de Séries Temporais e Econometria, 1997.
- Priestley, M. B. (1981), *Spectral Analysis and Time Series*. London: Academic Press.

Apêndice

Gráfico A.1: Coeficientes de Ondas Originals para a Série *ProdInd*:

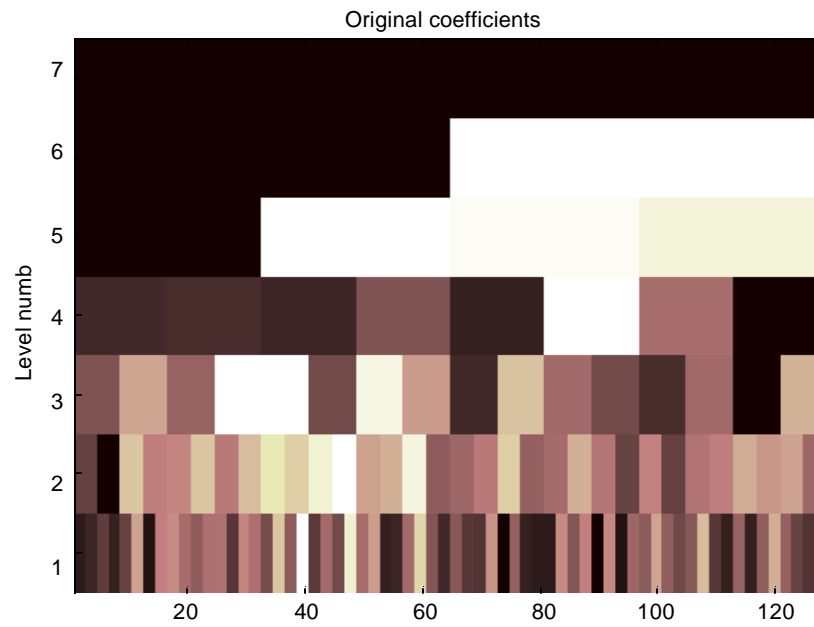


Gráfico A.2: Coeficientes de Ondas Resultantes do Processo de Retirada de Ruído para a Série *ProdInd*:

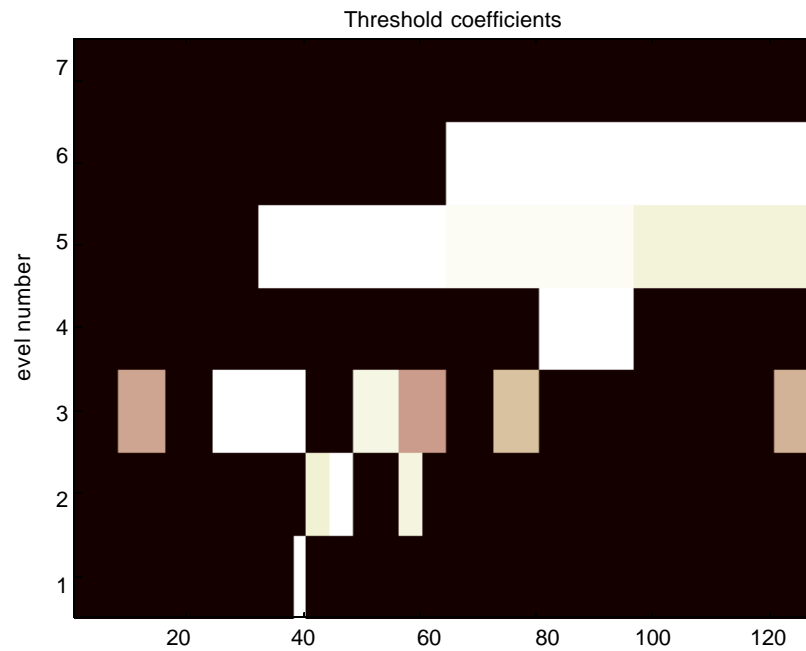


Gráfico A.3: Coeficientes de Ondaletas Originais para a Série *Expor*:

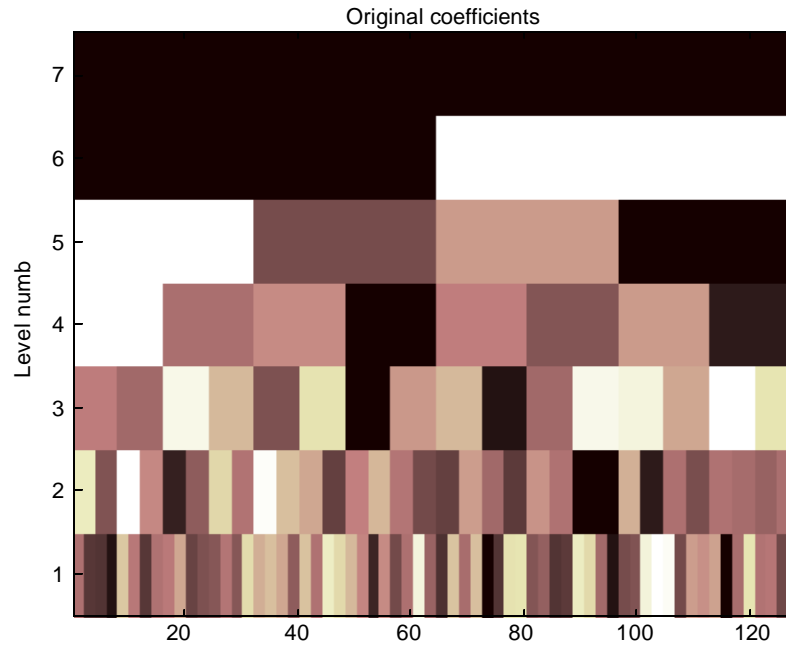


Gráfico A.4: Coeficientes de Ondaletas Resultantes do Processo de Retirada de Ruído para a Série *Expor*:

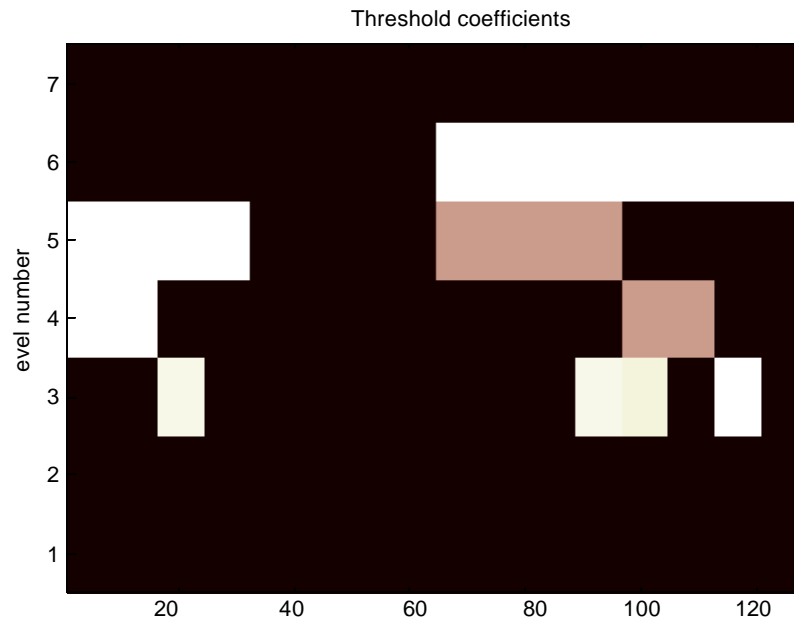


Gráfico A.5: Coeficientes de Ondas Originais para a Série *Fish*:

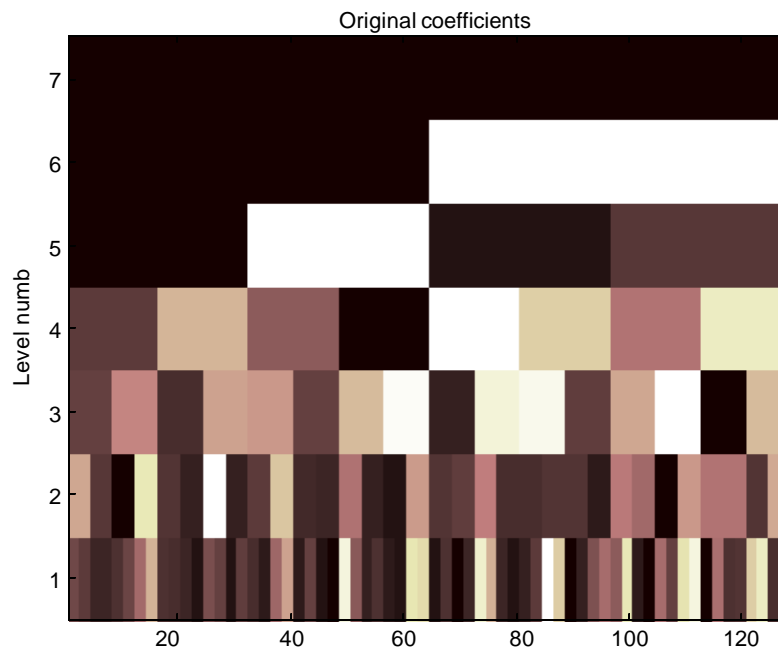


Gráfico A.6: Coeficientes de Ondas Resultantes do Processo de Retirada de Ruído para a Série *Fish*:

