

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA APLICADA

# **Wavelets: uma Aplicação à Estimação do Núcleo de Inflação Brasileiro**

por

Erik Stephanou Elsenbruch Filomena

Dissertação submetida como requisito parcial  
para a obtenção do grau de  
Mestre em Matemática Aplicada

Prof. Dr. Jean Carlo Pech de Moraes  
Orientador

Porto Alegre, Julho de 2018.

## CIP - CATALOGAÇÃO NA PUBLICAÇÃO

Stephanou Elsenbruch Filomena, Erik

Wavelets: uma Aplicação à Estimação do Núcleo de Inflação Brasileiro / Erik Stephanou Elsenbruch Filomena.— Porto Alegre: PPGMAp da UFRGS, 2018.

120 p.: il.

Dissertação (mestrado) —Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Matemática Aplicada, Porto Alegre, 2018.

Orientador: Pech de Moraes, Jean Carlo

Dissertação: Matemática Aplicada,  
Análise Aplicada

# Wavelets: uma Aplicação à Estimação do Núcleo de Inflação Brasileiro

por

Erik Stephanou Elsenbruch Filomena

Dissertação submetida ao Programa de Pós-Graduação em Matemática Aplicada do Instituto de Matemática da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do grau de

## Mestre em Matemática Aplicada

Linha de Pesquisa: Análise aplicada

Orientador: Prof. Dr. Jean Carlo Pech de Moraes

Banca examinadora:

Prof. Dr. Eduardo de Oliveira Horta  
Instituto de Matemática e Estatística  
Universidade Federal do Rio Grande do Sul

Prof. Dr. Lucas da Silva Oliveira  
Instituto de Matemática e Estatística  
Universidade Federal do Rio Grande do Sul

Prof. Dra. Maria Cristina Pereyra  
Department of Mathematics and Statistics  
University of New Mexico

Dissertação apresentada e aprovada em  
Julho de 2018.

Prof. Dr. Carlos Hoppen  
Coordenador

# Sumário

<b>LISTA DE FIGURAS</b> . . . . .	<b>vi</b>
<b>LISTA DE TABELAS</b> . . . . .	<b>vii</b>
<b>RESUMO</b> . . . . .	<b>viii</b>
<b>ABSTRACT</b> . . . . .	<b>ix</b>
<b>1 INTRODUÇÃO</b> . . . . .	<b>1</b>
<b>2 FILTROS LINEARES EM <math>\mathbb{R}^N</math></b> . . . . .	<b>4</b>
<b>2.1 Sinais em <math>\mathbb{R}^N</math></b> . . . . .	<b>5</b>
2.1.1 Convolução em $\mathbb{R}^N$ . . . . .	15
2.1.2 Correlação Cruzada . . . . .	29
<b>3 BANCOS DE FILTROS E A TRANSFORMADA WAVELET</b> . . . . .	<b>30</b>
<b>3.1 Wavelets como Transformadas Lineares</b> . . . . .	<b>32</b>
<b>3.2 Wavelets como Filtros Lineares</b> . . . . .	<b>47</b>
3.2.1 As Daubechies no domínio da frequência e os níveis superiores da DWT . . . . .	51
<b>3.3 A Maximum Overlap Discrete Wavelet Transform</b> . . . . .	<b>54</b>
3.3.1 Remoção de ruídos com wavelets . . . . .	58
3.3.1.1 Adaptação espacial com soft Thresholding. . . . .	64
<b>4 MEDINDO O NÚCLEO DE INFLAÇÃO</b> . . . . .	<b>67</b>
<b>4.1 Por que o núcleo de inflação.</b> . . . . .	<b>67</b>
4.1.1 Núcleo de inflação. . . . .	71
<b>4.2 Os Dados</b> . . . . .	<b>77</b>
4.2.1 O modelo . . . . .	79
<b>4.3 Medidas de Núcleo baseadas em Wavelets</b> . . . . .	<b>79</b>
<b>4.4 Medidas baseadas em wavelets com amostra restrita.</b> . . . . .	<b>91</b>
<b>5 CONCLUSÃO</b> . . . . .	<b>105</b>

REFERÊNCIAS BIBLIOGRÁFICAS . . . . .	109
6 APÊNDICE . . . . .	117

## Lista de Figuras

2.1	Polinômios amostrados a taxas $N$ diferentes . . . . .	7
2.2	Espectro de um vetor aleatório . . . . .	12
2.3	Espectro do sinal $h$ . . . . .	14
2.4	Função ganho quadrado da Wavelet de Haar . . . . .	24
2.5	Função ganho quadrado da Função Escala de Haar . . . . .	26
2.6	Sinal $x(t)$ filtrado. . . . .	28
3.1	Exemplo DB2 . . . . .	45
3.2	Função ganho quadrado da função escala DB2 . . . . .	52
3.3	Função ganho quadrado da wavelet DB2 . . . . .	53
4.1	Tempo de vigência em dias da taxa meta dos <i>Federal Funds</i> . . . . .	68
4.2	Tempo de vigência em dias da taxa meta SELIC . . . . .	69
4.3	CPI mensal à taxas anualizadas entre dez/1995 e dez/2017 . . . . .	70
4.4	IPCA mensal a taxas anualizadas entre dez/1995 e mar/2018 . . . . .	77
4.5	IPCA Estetendido por Reflexão . . . . .	82
4.6	IPCA e Núcleos de Inflação . . . . .	84
4.7	Coeficientes do primeiro nível <i>detail</i> da transformada wavelet . . . . .	92
4.8	Coeficientes do segundo nível <i>detail</i> da transformada wavelet . . . . .	93
4.9	IPCA e núcleos WIM com amostra restrita - Núcleos baseados em wavelets . . . . .	97
4.10	IPCA e núcleos WIM com amostra restrita - Núcleos oficiais . . . . .	98
6.1	Tabela dos itens que compõe o núcleo por exclusão 2 (EX2) . . . . .	117
6.2	Scatter Plot Inflação/Núcleo . . . . .	118
6.3	<i>Scatter plot</i> IPCA/núcleo com amostra restrita . . . . .	119
6.4	<i>Scatter plot</i> IPCA/núcleo com amostra restrita . . . . .	120

## Lista de Tabelas

4.1	Estatísticas descritivas . . . . .	83
4.2	Teste de viés . . . . .	86
4.3	Teste de Wald para restrição nos coeficientes . . . . .	87
4.4	Teste de ajuste dinâmico . . . . .	88
4.5	Teste de poder preditivo dos núcleos . . . . .	91
4.6	Estatísticas descritivas e teste de normalidade para amostra reduzida	99
4.7	Testes de estacionariedade . . . . .	99
4.8	Teste de viés para amostra reduzida . . . . .	100
4.9	Teste de Wald para restrição nos coeficientes com amostra restrita . .	101
4.10	Teste de ajuste dinâmico com amostra restrita . . . . .	101
4.11	Teste de poder preditivo dos núcleos . . . . .	103
4.12	Comparação de RMSE . . . . .	104

# RESUMO

Wavelets são descritas como sendo capazes de dar tanto resolução em frequência como resolução temporal a um sinal. Este trabalho revisa o que é o domínio da frequência em um espaço de dimensão finita, como o  $\mathbb{R}^N$  e apresenta como a Transformada Wavelet Discreta e a *Maximum Overlap Discrete Wavelet Transform* podem ser usadas para decompor um sinal em diversos componentes de escalas, que podem ser vistos como componentes de frequência ou componentes temporais. Então uma aplicação para a estimação do núcleo de inflação para o IPCA oficial brasileiro é apresentada. Ela consiste em obter uma Análise Multirresolução baseada na wavelet Daubechies 2 e estimar a inflação subjacente, ou removendo-se níveis *detail*, ou aplicando um algoritmo de *threshold*. Por último, alguns testes de qualidade de medida sugeridos pela literatura são executados. Isso é feito com o conjunto completo dos dados e com um conjunto restrito, obtido com um método baseado em wavelets para detecção de quebras estruturais em séries temporais.



# ABSTRACT

Wavelets are described as being able to give both a time resolution and a frequency resolution to a signal. This work reviews what is the frequency domain when represented by a finite dimensional space such as the  $\mathbb{R}^N$  and presents how the Discrete Wavelet Transform and the Maximum Overlap Discrete Wavelet Transform can be used to decompose a signal in several scale components, which can be viewed as frequency components or as time components. Then an application to the estimation of the core inflation for the official Brazilian CPI is presented. This is done by obtaining a Multi Resolution Analysis based on the Daubechies 2 wavelet and estimating the underlying inflation rate by either removing detail levels completely or applying a threshold algorithm. Lastly, a few tests of quality of measurement proposed by the literature are performed. This is done with the full set of data and a restricted set, obtained with a wavelet method for detecting structural breaks in time series.

# DECLARAÇÃO

O presente trabalho foi realizado com apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil

# 1 INTRODUÇÃO

A utilização de técnicas do domínio da frequência (i.e. uso de técnicas de análise de componentes periódicos) para modelar e analisar problemas econômicos não é novidade, mesmo que, como menciona Steehouwer [56], ela não seja reconhecida como tal. As séries temporais, como o próprio nome diz, possuem uma identificação com o tempo, mas além disso, podemos estudar a forma como elas parecem se repetir, ou oscilar, com certa harmonia, isto é, a sua representação no domínio da frequência. Quando econométricos tomam diferenças de séries temporais com o intuito de torná-las estacionárias, eles estão aplicando um filtro às séries, que como será visto, é chamado Filtro de Haar. Uma introdução à aplicação de ferramentas do domínio da frequência em economia é dada por Steehouwer [56]. Em particular este trabalho está interessado na discretização do domínio da frequência.

Wavelets entram nesse contexto como filtros. Neste caso, filtros são transformadas lineares que agem no domínio da frequência selecionando certas frequências e alterando suas amplitudes e fases. Assim é possível, por exemplo, remover ruídos de um sinal, desde que estes sejam compostos por frequências diferentes. Além disso, wavelets também são descritas por terem, além de resolução em frequência, resolução espacial, onde por espacial entende-se algo que possua dimensão (em particular, o tempo).

A vantagem de se ter essa resolução espacial é o ganho de flexibilidade. No domínio da frequência, as coisas são rígidas: se uma determinada frequência está presente em uma parte do sinal, então ela está presente em todo sinal. E ela está presente sempre com a mesma amplitude. Isso não permite detectar mudanças no sinal (imagine mudanças de música em uma rádio). Por outro lado wavelets, além das propriedades no domínio da frequência, são descritas por Schleicher [51, p.1, tradução do autor.] como "[...] lentes de uma câmera que permitem alguém tirar fotos de uma ampla paisagem bem como dar *zoom* em um detalhe microscópico que normalmente é invisível ao olho humano.". Isso nos dá flexibilidade de escolher como modelar um sinal; se usamos informação localizada (tempo) ou informação geral (frequência). Para executar esta tarefa com sinais discretos utilizaremos a Transformada Wavelet Discreta (*Discrete Wavelet Transform*, ou DWT).

Entre as variáveis econômicas que podem ser estudadas como um sinal está a inflação, a qual pode ser vista como um sinal poluído por um ruído. Em especial, é de interesse remover o ruído deste sinal, de forma a recuperar o chamado núcleo de inflação, o qual representa a tendência subjacente de aumento nos preços, livre de oscilações de curto prazo. O interesse dos agentes econômicos em descobrir o núcleo é que, em teoria, apenas este deve ser alvo da política monetária, enquanto que efeitos transitórios podem ser ignorados. Um breve histórico deste conceito pode ser encontrado em Wynne [63]. Contudo, não é claro o que representa o sinal de fato e o que representa ruído. Por isso na literatura várias medidas de núcleo são propostas, e em geral se aceita que uma boa medida deve ser uma medida não enviesada da inflação subjacente e que seja capaz de ajudar a prever a inflação futura.

Neste trabalho serão apresentadas as duas formas de se interpretar a transformada Wavelet. O segundo capítulo vai apresentar a modelagem do domínio da frequência em um espaço de dimensão finita. Os vetores serão interpretados como observações de um sinal e, para estudá-los, será visto como as funções sinusoides discretizadas formam uma base para o espaço em questão. Associadas a esta base estão as matrizes circulantes e com elas será visto o Teorema da Convolução o qual permite, entre outras coisas, estudar as Transformadas Wavelet no domínio da frequência.

O terceiro capítulo usa a teoria geral da álgebra linear e apresenta a DWT como uma transformada linearlinear, de fato uma transformação linear ortogonal, logo invertível, de forma que as colunas de sua matriz formam uma base de  $\mathbb{R}^N$ . Será visto como esta base dá a interpretação temporal para as wavelets discretas. Em seguida serão aplicadas as ferramentas vistas no capítulo 2 para ver como as wavelets também possuem resolução em frequência, e como esta resolução permite obter com facilidade uma extensão da DWT chamada *Maximum Overlap Discrete Wavelet Transform* (MODWT). Enquanto que a DWT é ineficiente quando um sinal é amostrado em datas diferentes, a MODWT não sofre deste problema, e é descrita como uma boa alternativa para utilização de wavelets com séries temporais. No fim deste capítulo será vista uma técnica de estimação de sinais, desenvolvida por Donoho e Johnstone [17], chamada de *thresholding*, baseada nos aspectos temporais da DWT e da MODWT e que permite estimar com mais precisão um sinal que é poluído por ruídos.

No quarto capítulo serão utilizadas as ferramentas vistas nos capítulos 2 e 3 no problema de obtenção do núcleo de inflação para o caso brasileiro. A primeira utilização de wavelets neste problema foi proposta por Dowd e Cotter [18] e feita para o Brasil primeiramente por Denardin, Kozakevicius e Schmidt [16]. No presente trabalho avançamos em relação ao último trabalho com a inclusão de medidas de núcleo com a etapa de *thresholding*. Aqui a inflação é interpretada como um sinal poluído por ruídos e é utilizada a análise multirresolução com wavelets para extrair o sinal dos dados empíricos. Para avaliar se as estimativas de núcleo são adequadas serão feitos testes sugeridos na literatura que confirmam se um candidato satisfaz os requerimentos teóricos para ser uma medida de núcleo. Para fins de comparação, junto das medidas baseadas em wavelets serão apresentadas e testadas as medidas oficiais calculadas pelo Banco Central do Brasil.

## 2 FILTROS LINEARES EM $\mathbb{R}^N$

No livro *Théorie analytique de la chaleur*, publicado em 1822 [28] Joseph Fourier (1768-1830) estudou a condução do calor em meios sólidos e descreveu a função matemática que rege sua difusão por meio de séries periódicas infinitas, conhecidas hoje como séries de Fourier [57]. Essas séries são combinações lineares de funções senos e cossenos, avaliadas nas frequências inteiras positivas. Sob algumas hipóteses sobre a imagem da função, pode-se escrevê-la como:

$$f(x) = \sum_{t=0}^{+\infty} a_t \cos(2\pi xt) + b_t \sin(2\pi xt).$$

A representação de funções em séries de Fourier introduz a ideia de domínio da frequência para representar objetos que são observados em outros domínios, por exemplo o tempo. Um fato importante sobre a série de Fourier como mostrada é que ela é periódica na reta, com período unitário. Assim se  $h$  for inteiro, então  $f(x+h) = f(x)$ . Isso limita o uso de séries de Fourier a funções periódicas ou com suporte em um intervalo limitado, pois sempre pode-se estender a série periodicamente fora dele. Por outro lado se  $f$  não for periódica, mas for integrável, ainda é possível escrevê-la no domínio da frequência por meio do que se chama de transformada de Fourier [14], dada por:

$$f(x) = \int_{-\infty}^{+\infty} F(\omega) e^{2\pi i \omega x} d\omega \quad (2.1)$$

onde  $\omega$  representa o contínuo das frequências e por razões estéticas usa-se a fórmula de Euler para representar senos e cossenos. A função  $F$  é obtida de  $f$  por meio da transformada de Fourier:

$$F(\omega) = \int_{-\infty}^{+\infty} f(x) e^{2\pi i \omega x} dx. \quad (2.2)$$

Quando se fala em representação no domínio da frequência, o que se deseja saber é quais são os coeficientes  $a_t$  e  $b_t$ , ou qual é a função  $F$  que satisfazem a igualdade. Caso eles sejam conhecidos, então saberemos a representação no domínio da frequência de uma função. Por outro lado, dados os coeficientes, podemos representar - no domínio do tempo - algo que está dado no domínio da frequência. Será utilizada a notação de Persival e Walden [46] para representar essa dualidade:  $f \longleftrightarrow (a_t, b_t)$  ou  $f \longleftrightarrow F$ . A relação entre as duas funções satisfaz a condição para ser um isomorfismo (leva somas à somas e

produtos por escalar a produtos por escalar) [41]. Um dos problemas da análise de Fourier é descobrir em que casos este procedimento pode ser feito e como achar os coeficientes de representação de  $f$ . Neste trabalho será utilizado um caso especial de transformada de Fourier, a transformada discreta de Fourier (DTF, do acrônimo em inglês) *Discrete Time Fourier Transform*. Sucintamente, a análise feita com a DFT reescreve vetores em  $\mathbb{C}^N$  em uma base cujos elementos são obtidos mediante discretização das funções seno e cosseno. Para isso Broughton e Bryan (2008) definem os vetores  $E_k$ :

$$E_k = [e^{\frac{2\pi ikt}{N}}]_{t=0}^{N-1}. \quad (2.3)$$

Aqui devemos notar o fato de que em (2.3) estamos lidando com vetores complexos. Usamos o isomorfismo entre espaços de dimensão finita para poder utilizar uma base complexa para expressar vetores reais. As entradas do vetor  $E_k$  são os valores das funções sinusoides de frequência  $\frac{k}{N}$  [46] avaliadas no instante  $t$ . Lembrando que o produto escalar complexo  $\langle x, y \rangle$  é dado por  $\sum_{t=0}^{N-1} x_t \bar{y}_t$ , definimos os coeficientes de Fourier do vetor  $x \in \mathbb{C}^N$  como:

$$\mathbf{X}_k = \sum_{t=0}^{N-1} x_t e^{-\frac{2\pi ikt}{N}}. \quad (2.4)$$

A motivação do uso da transformada discreta de Fourier ao invés da transformada contínua é quando só se é possível obter valores amostrados das funções que se deseja estudar em um *grid* ( $\{t_0, t_1, \dots, t_{N-1}\}$ ). Além disso, essa simplificação facilita o estudo no domínio da frequência pois pode-se usar as ferramentas usuais da álgebra linear. Assim, Percival e Walden [46], Broughton e Bryan [9] e Walker [58] fazem quase exclusivamente o uso de álgebra linear para desenvolver a teoria dos filtros de dados e das transformadas Wavelet.

## 2.1 Sinais em $\mathbb{R}^N$

Os objetos de estudo deste capítulo serão sinais obtidos ao amostrar-se uma função com valores reais durante um intervalo de tempo. Por sinal entende-se alguma informação que é emitida por uma fonte emissora e recebida por algum instrumento de captação, que faz observações sequenciadas de acordo com o tempo. Para esta modelagem inicial será utilizada a abordagem de Broughton e Bryan [9].

Suponha que em um intervalo de tempo qualquer, o qual é normalizado, sem perda de generalidade, para  $[0, 1]$ , sejam observados  $N$  valores atribuídos a um fenômeno. O intervalo entre cada observação, chamado taxa de amostragem, é igual a  $\frac{1}{N}$ . Estes valores podem ser escritos em uma lista  $(x_0, x_1, x_2, \dots, x_{N-1})$ , e fica claro que uma boa representação matemática para as observações obtidas são vetores em  $\mathbb{R}^N$ . Podemos então usar a estrutura deste espaço vetorial para modelar o sinal; em especial podemos somar sinais, multiplicá-los por escalares e reescrevê-los em bases diferentes. No entanto, existem algumas preocupações quando se modela um sinal desta forma. Boughton e Bryan [9] citam três fontes de problemas. A primeira diz respeito à amostragem. Quando se amostra um sinal de forma discreta é bem provável que ocorra perda de informação durante o processo, pois como *a priori* não se conhece a função emissora, não se saberá ao certo se a frequência de observações é adequada ou não. Um exemplo disso é imaginar dois polinômios  $p_1$  e  $p_2$  que se interceptam em exatamente  $N$  pontos. Suponha que uma fonte emita um sinal com valores iguais a um deles. Se forem realizadas observações exatamente nos pontos de intercepto, seria impossível saber qual deles que se está medindo. Por outro lado se realizássemos  $k > N$  observações, já seria possível identificar o sinal subjacente. A questão da amostragem é similar ao problema dos polinômios.

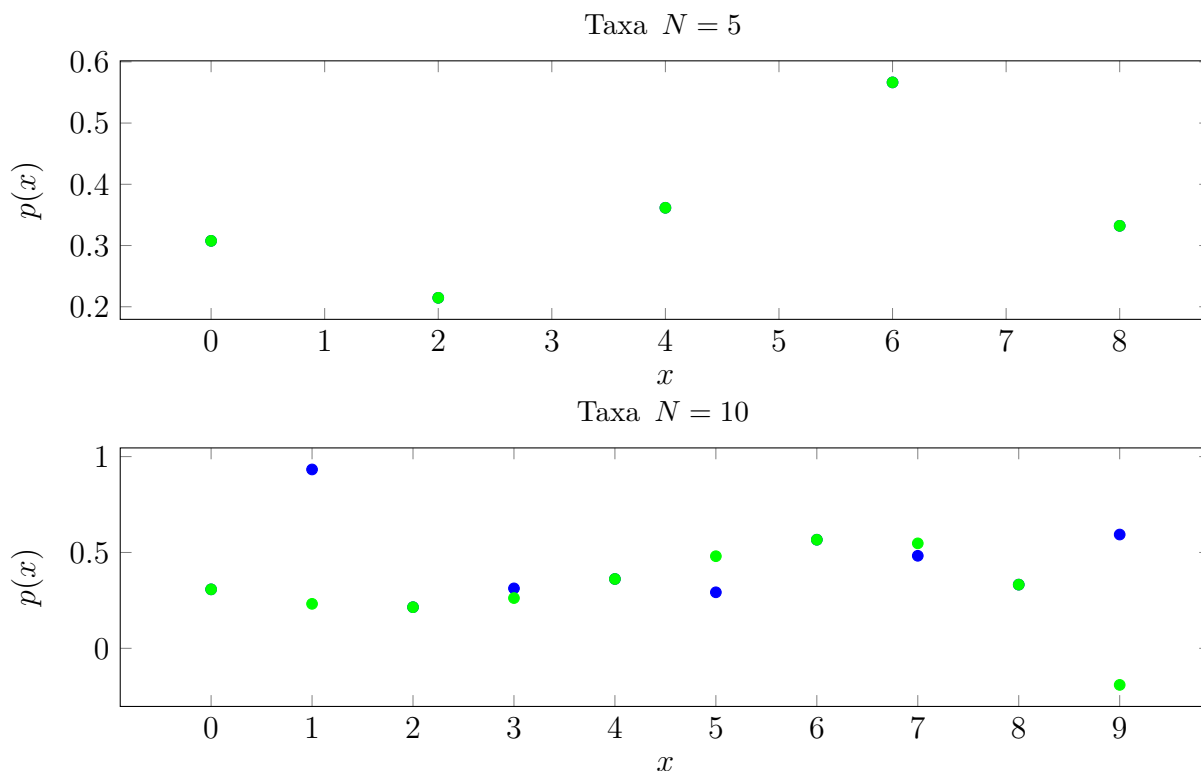
Na figura 2.1 são plotados os pontos observados de dois polinômios de graus quatro e nove, respectivamente, gerados aleatoriamente, mas que coincidem em cinco pontos. Se forem realizadas cinco amostras no intervalo  $[0, 9]$  e por algum azar elas forem feitas nos instantes que eles se interceptam, é impossível distinguir os valores, mas se forem realizadas dez amostras, em quaisquer pontos, já é possível diferenciá-los.

A segunda fonte de problema é a quantização inerente ao procedimento de observação e armazenamento do sinal, também chamado de **erro de arredondamento**. Ao utilizar aritmética de ponto flutuante perde-se alguma informação. Espera-se que com aproximações maiores (i.e. aumento do número de casas decimais) o problema seja reduzido, mas sempre existe a possibilidade de erro [9]. Como neste trabalho serão utilizados dados provenientes de fontes externas, este erro de medição não será considerado.

A terceira fonte de problemas é devida ao sinal em si. A fonte emissora transmite um sinal de acordo com algum processo, o qual pode ser tanto uma função determinística quanto a realização de um processo estocástico. Em qualquer caso, o sinal de interesse



**Figura 2.1:** Polinômios amostrados a taxas  $N$  diferentes



Fonte: Elaborado pelo autor, baseado em Broughton e Bryan [9]

usualmente é observado com **ruído**, às vezes chamado de erro.<sup>1</sup>, às vezes chamado de erro. Aqui será assumido que o erro é aditivo e pode ser modelado como um sinal em si. Desta forma se decompõe o sinal da seguinte maneira [9]:

$$x_t = \phi_t + \epsilon_t \tag{2.5}$$

Onde  $x_t$  representa o sinal observado no tempo  $t$ ,  $\phi_t$  o sinal de interesse e  $\epsilon_t$  um termo de ruído sobre o qual serão feitas hipóteses mais adiante. Denotamos a **observação do sinal no instante  $t$**  por  $x_t$ , e por  $x := (x_0, x_1, \dots, x_{N-1})$  a **amostra** do sinal. Para desenvolver a teoria dos filtros não é necessário assumir a existência do termo de ruído então por ora assume-se que é um vetor nulo. Sucintamente, quando se aplica um filtro em um sinal, tanto discreto como contínuo, estamos agindo em uma base do espaço ao qual este sinal pertence e modificando os coeficientes da combinação linear entre os vetores

---

<sup>1</sup>Em econometria chama-se o ruído de erro de previsão, ou em modelos de crescimento econômico, de medida de nossa ignorância [39]. Neste trabalho serão utilizados os termos erro ou ruído para significar a parte do sinal que não representa o sinal em si mas variações decorrentes de fontes externas.

que o geram. Voltando a (2.3), se construirmos os vetores  $E_k$  para  $k$  inteiro, então esse conjunto de  $N$  vetores será uma base para  $\mathbb{C}^N$ . A prova, seguindo Broughton e Bryan [9, p. 35] é dada a seguir. Ela usa a seguinte identidade:

$$1 + z + z^2 + \dots + z^{N-1} = \frac{1 - z^N}{1 - z}. \quad (2.6)$$

Para  $k \neq l$ ,  $k, l \in \left(-\frac{N}{2}, \frac{N}{2}\right]$  temos

$$\begin{aligned} \langle E_k, E_l \rangle &= \sum_{r=0}^{N-1} e^{2\pi ikr/N} \overline{e^{2\pi ilr/N}} \\ &= \sum_{r=0}^{N-1} e^{2\pi ikr/N} e^{-2\pi ilr/N} \\ &= \sum_{r=0}^{N-1} (e^{2\pi i(k-l)/N})^r. \end{aligned} \quad (2.7)$$

Utilizando a identidade (2.6) com  $z = e^{2\pi i(k-l)r/N}$  temos

$$\begin{aligned} \langle E_k, E_l \rangle &= \frac{1 - (e^{2\pi i(k-l)/N})^N}{1 - e^{2\pi i(k-l)/N}} \\ &= \frac{1 - e^{2\pi i(k-l)}}{1 - e^{2\pi i(k-l)/N}} \\ &= 0, \end{aligned}$$

pois  $1 - e^{2\pi i(k-l)} = 0$  para quaisquer  $k \neq l$  inteiros e o denominador nunca se anula no intervalo. Portanto os  $N$  vetores  $E_k$  são ortogonais entre si e formam uma base para  $\mathbb{C}^N$ . Logo, podemos reescrever qualquer vetor  $x$  de forma única como

$$x = \sum_{i=0}^{N-1} a_i E_i.$$

Já que esta base é ortogonal, onde os termos  $a_i$  têm a forma

$$a_i = \frac{\langle x, E_i \rangle}{\langle E_i, E_i \rangle}.$$

Lembrando da identidade trigonométrica  $\cos^2(x) + \sin^2(x) = 1$ , junto com a regra do produto interno para vetores complexos, temos  $\langle E_i, E_i \rangle = N$ , e conseqüentemente,

$$x = \frac{1}{N} \sum_{i=0}^{N-1} \langle x, E_i \rangle E_i$$

o qual comparamos com (2.4) para obter

$$x = \frac{1}{N} \sum_{i=0}^{N-1} \mathbf{X}_i E_i. \quad (2.8)$$

Este resultado garante que existe uma base para todos os sinais amostrados em um dado *grid* finito de pontos  $\{t_0, t_1, \dots, t_{N-1}\}$ , e sem perda de generalidade consideramos  $t_k = \frac{k}{N}$ . As equações (2.4) e (2.8) definem a transformada discreta de Fourier (DFT) e a transformada inversa de Fourier (IDFT) para vetores em  $\mathbb{C}^N$  [9]. Elas são as análogas discretas de (2.1) e (2.2). Em especial, podemos usar o fato de que cada vetor complexo sem parte imaginária representa um único vetor real puro e usar a base de Fourier para estudar  $\mathbb{R}^N$ . Existe um fato importante sobre os coeficientes de Fourier de um vetor real. Seguindo Broughton e Bryan [9]:

$$\begin{aligned} \mathbf{X}_{-k} &= \sum_{t=0}^{N-1} x_t e^{\frac{-2\pi i k t}{N}} \\ &= \sum_{t=0}^{N-1} x_t e^{\frac{2\pi i k t}{N}} \\ &= \sum_{t=0}^{N-1} \overline{x_t e^{\frac{-2\pi i k t}{N}}} \\ &= \overline{\sum_{t=0}^{N-1} x_t e^{\frac{-2\pi i k t}{N}}} \\ &= \overline{\mathbf{X}_k} \end{aligned} \quad (2.9)$$

Alternativamente, se um sinal tem a DFT tal que  $\mathbf{X}_k = \overline{\mathbf{X}_{-k}}$  então este sinal só terá componentes reais [9]. O foco do restante deste trabalho serão estes vetores.

Para aplicar a base de Fourier à teoria dos filtros, são feitas algumas observações. A primeira delas é que cada  $E_k$  corresponde a uma função senoide de uma certa frequência, pois cada vetor  $E_k$  ( $k$  fixado) corresponde aos valores da função

$$t \mapsto \cos\left(\frac{2\pi k t}{N}\right) + i \operatorname{sen}\left(\frac{2\pi k t}{N}\right)$$

avaliada nos pontos  $t \in \{0, 1, \dots, N-1\}$ . Esta função é periódica com período  $\frac{N}{k}$  e frequência  $\frac{k}{N}$ . Para simplificar a notação dizemos que a frequência é simplesmente  $k$ , isto é, em um intervalo de tamanho 1 a função executa  $k$  ciclos completos, se  $N$  for par, ou  $k/N$

ciclos completos se  $N$  for ímpar. A frequência está relacionada com a taxa de amostragem, não com o tamanho do intervalo, portanto falar em uma frequência  $k$  significa que em um instante  $1/N$  foi executado  $k/N$  de um ciclo. Também nota-se que se trocássemos  $k$  por  $k + N$  a função teria os mesmos valores para  $t \in \{0, 1, 2, \dots, N - 1\}$ . Isso quer dizer que se estivéssemos amostrando diversas funções sinusoides com uma taxa de amostragem de  $\frac{1}{N}$  seria impossível distinguir entre os valores obtidos de uma senoide e os valores obtidos da senoide com a frequência deslocada em  $N$  unidades. Esse fenômeno é o problema da amostragem apontado por Broughton e Bryan [9]. Assim como existe incerteza quanto a qual polinômio estamos observando, existe incerteza sobre qual a frequência real que observamos. No caso da análise de frequências, chama-se esse fenômeno de *aliasing*.

**Definição 1.** Os vetores  $E_k$ , para  $k = 0, 1, 2, \dots, N - 1$  e com entradas dadas pelas funções sinusoides  $e^{\frac{2\pi i k t}{N}}$  para  $t = 0, 1, 2, \dots, N - 1$  formam uma base ortogonal para  $\mathbb{C}^N$  chamada de **Base de Fourier**.

Como consequência do *aliasing* temos o Teorema da Amostragem de Shannon-Nyquist, que afirma que para captar frequências de magnitude  $W$  ciclos por segundo é necessário amostrar um sinal a uma taxa amostral de  $\frac{1}{2W}$  [52], com o intervalo de tempo entre amostras fixo. Mais forte do que isso, o teorema de Shannon-Nyquist afirma que se um sinal tem como componentes frequências de no máximo  $W$ , então com  $2W$  observações a informação sobre ele pode ser completamente recuperada. Chama-se de **frequência de Nyquist** aquela frequência que representa meio ciclo dado a cada intervalo de observação, ou seja, a frequência máxima que podemos recuperar de um sinal a uma dada taxa de amostragem. A frequência de Nyquist será denotada  $f_N$ .

Uma observação que deve ser feita é que utiliza-se frequências negativas quando trabalha-se com análise de Fourier. Esta é apenas uma convenção matemática [46], adotada por diversos autores, já que - devido ao *aliasing* - qualquer intervalo de medida 1 vai conter a mesma informação sobre as frequências. Da maneira como foi definida, a transformada de Fourier só considera as frequências positivas, mas na literatura em geral são encontradas frequências negativas. Para entender o porquê é só notar que se

$\frac{N}{2} \leq k \leq N$  então  $E_k = E_{k-N}$ . Isso implica que para provar (2.7) tanto faz  $k$  estar no intervalo  $[0, N-1]$  ou  $\left[\frac{-N}{2}, \frac{N-1}{2}\right]$ .

Usando a notação mencionada para relacionar uma função com sua série de Fourier, a relação de  $x$  e sua transformada discreta de Fourier é dada por  $\{x\} \longleftrightarrow \{\mathbf{X}\}$ . Serão utilizadas as seguintes propriedades dos coeficientes  $\mathbf{X}_k$ :

**Proposição 1.**

- (i)  $\mathbf{X}_0 = \sum_{i=0}^{N-1} x_i$  ou  $\frac{\mathbf{X}_0}{N} = \bar{x}$ , onde  $\bar{x}$  é a média temporal dos valores de  $x$ .
- (ii) **Teorema de Parseval** [46]:  $\|x\|_2 = \frac{1}{N^{\frac{1}{2}}} \|\mathbf{X}\|_2$ .

*Demonstração.* Para provar (i) basta notar que  $E_0$  é um vetor de uns. Para provar (ii), utiliza-se (2.8) e toma-se o produto interno de cada lado por si mesmo:

$$\begin{aligned} \langle x, x \rangle &= \left\langle \frac{1}{N} \sum_{i=0}^{N-1} \mathbf{X}_i E_i, \frac{1}{N} \sum_{i=0}^{N-1} \mathbf{X}_i E_i \right\rangle \\ &= \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} \mathbf{X}_i \bar{\mathbf{X}}_k \langle E_i, E_k \rangle \\ &= \frac{1}{N^2} \sum_{k=0}^{N-1} \mathbf{X}_x \bar{\mathbf{X}}_k N \\ &= \frac{1}{N} \sum_{k=0}^{N-1} |\mathbf{X}_k|^2. \end{aligned}$$

Tomando as raízes de ambos os lados, tem-se a igualdade. Salienta-se que o Teorema de Parseval usa apenas que a base é composta por vetores ortogonais. Esse resultado serve para qualquer base ortogonal em espaços de Hilbert.  $\square$

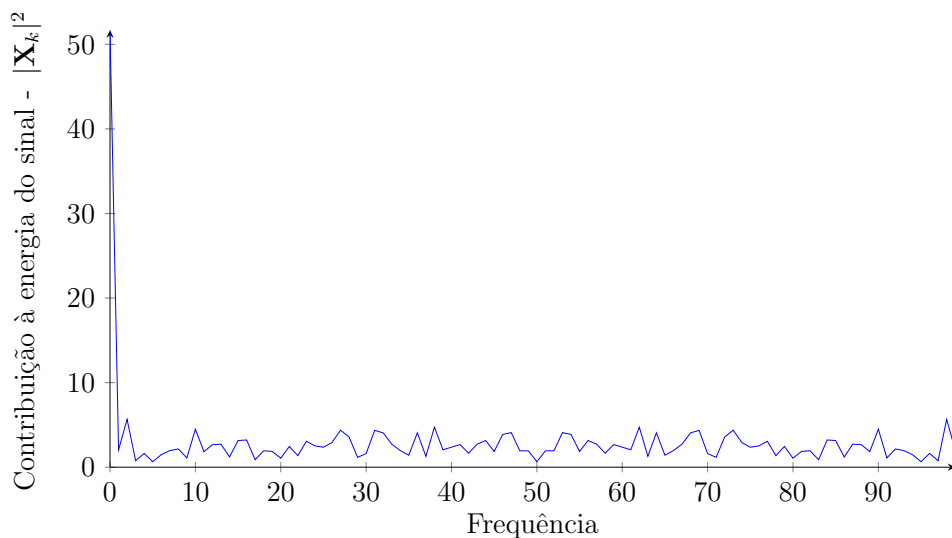
Uma consequência do Teorema de Parseval é que pode-se decompor a norma de  $x$  em frequências e estudar o quanto cada uma delas contribui para a energia<sup>2</sup> do

---

<sup>2</sup>Persival e Walden [46] observam que o termo energia é uma nomenclatura para se referir à norma do vetor e não guarda necessariamente relação com energia no sentido físico

sinal, visto que a DFT preserva (exceto pela multiplicação de uma constante conhecida) a norma do vetor  $x$ . Seguindo Broughton e Bryan [9] monta-se um gráfico chamado espectrômetro que plota  $|\mathbf{X}_k|^2$  em função de  $k$ . A figura 2.2 mostra o espectro de um vetor aleatório gerado com o comando `rand` do Scilab<sup>3</sup>. Em um caso real, ele pode ser utilizado para mostrar quais frequências têm a maior participação na energia de um vetor. Neste caso a frequência constante<sup>4</sup> dá a maior contribuição ao sinal, enquanto as outras dão contribuições menores. A energia do sinal pode ser vista como uma medida

**Figura 2.2:** Espectro de um vetor aleatório



Fonte: elaborado pelo autor.

de variabilidade, já que a norma euclidiana de um vetor está relacionada ao segundo momento não centralizado da variável observada. A ideia por trás do espectro é escolher uma coleção  $J$  qualquer dos coeficientes  $\mathbf{X}_k$  que certamente terá a soma (dividida por  $N$ ) dos módulos menor que  $\|x\|^2$ , mas para um valor de tolerância  $\varepsilon > 0$ ,

$$\|x\|_2^2 - \varepsilon \leq \frac{1}{N} \sum_{k \in J} |\mathbf{X}_k|^2. \quad (2.10)$$

<sup>3</sup>O *default* deste comando gera valores uniformemente distribuídos no intervalo  $[0, 1]$

<sup>4</sup>O vetor obtido ao amostrar-se  $e^0$  será chamado de frequência constante. Outro nome comum na literatura é frequência dc (*direct current*).

Escrevendo  $x$  como (2.8), omitindo a divisão por  $N$ , subtraindo os coeficientes em  $J$  e tomando a norma

$$\left\| \sum_{k=0}^{N-1} |\mathbf{X}_k|^2 - \sum_{k \in J} |\mathbf{X}_k|^2 \right\|_2 \leq \varepsilon,$$

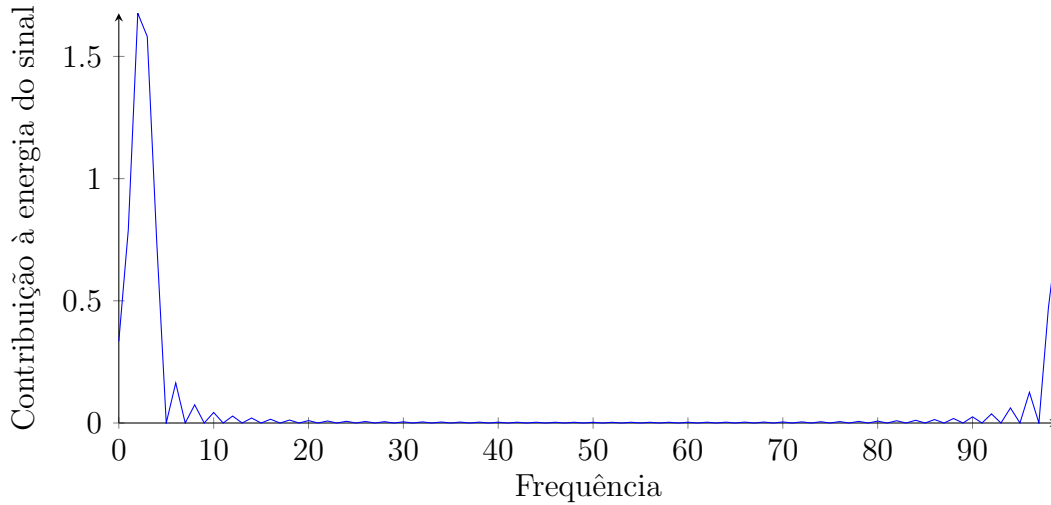
isto é, após remover certas frequências, o vetor obtido continua a no máximo  $\varepsilon$  de distância de  $x$ , na métrica induzida pela norma euclidiana. Esse procedimento de remover frequências mantendo proximidade à norma do vetor vai ser estudado adiante, pois ele permite, entre outras coisas, remover os ruídos em um sinal dado por (2.5).

Outra coisa importante pode ser notada no espectrômetro apresentado. Exatamente na coordenada abscissa 50, o gráfico se transforma num espelho, ou seja, ele repete os valores anteriores, na ordem inversa. Isso se dá porque sempre que um sinal é real, devido ao *aliasing* das frequência positivas e negativas, as frequências maiores do que  $N/2$  representam as frequências negativas entre  $-1/2$  e 0 [9]. Pelo Teorema de Shannon-Nyquist não é possível diferenciar frequências positivas maiores que a frequência de Nyquist das frequências negativas maiores do que  $-1/2$ . Por isso em geral é apresentado apenas metade dos coeficientes do espectro, já que a outra vai ser sempre igual, exceto por uma conjugação. Caso o sinal seja complexo isto não acontecerá.

Implicitamente, quando um sinal é analisado com a transformada discreta de Fourier assume-se que ou ele é periódico (afinal se forem estendidos os vetores da base, suas entradas vão se repetir periodicamente) ou existem mais informações sobre o sinal que podemos conseguir se for aumentada a taxa de amostragem. Também é assumido que todas as frequências têm participação igual durante todo o tempo no qual o sinal é observado, isto é, se um componente de frequência  $f_j$  aparece em algum momento, então ele vai continuar aparecendo indefinidamente. Considere o seguinte exemplo, baseado em Broughton e Bryan [9, p. 182-183]. Sejam dois sinais,  $x_t = \cos(2\pi t)$  e  $y_t = e^{\frac{i6\pi t}{100}}$ , definidos no intervalo  $[0, 1]$  e amostrados a uma taxa  $\frac{1}{100}$ . Aplicando a DFT e a IDFT em ambos obtemos a seguinte relação:

$$\begin{aligned} x &= \frac{1}{2}E_1 + \frac{1}{2}E_{99} \\ y &= E_3 \end{aligned}$$

**Figura 2.3:** Espectro do sinal  $h$



Fonte: elaborado pelo autor. Os valores do espectro foram calculados e depois tomada a raiz para ajustar a escala.

Note que  $x$  tem dois coeficientes de Fourier e que  $\mathbf{X}_1 = \mathbf{X}_{99}$ , pois  $x$  é um sinal real puro periódico de frequência 1, enquanto  $y$  é um sinal imaginário, e de fato, é um dos vetores da base de Fourier, para  $k = 3$ . Por outro lado, se definíssemos o sinal  $h_t$  da seguinte forma

$$h(t) = \begin{cases} x_t, & \text{se } 0 \leq t < \frac{1}{2} \\ y_t, & \text{se } \frac{1}{2} \leq t < 1. \end{cases}$$

Obteríamos um sinal composto por duas frequências que aparecem isoladamente em intervalos distintos. A figura 2.3 mostra o espectro de  $h$  e como se pode ver, mesmo que o sinal original tenha só três frequências, sua transformada de Fourier tem coeficientes não nulos em várias outras frequências. Isto ocorre por causa da descontinuidade do sinal em  $t = \frac{1}{2}$  e da incapacidade da transformada de Fourier de captar eficientemente descontinuidades no sinal.

Tendo em vista a dificuldade apontada acima, entre outros motivos, foram introduzidas alternativas à Transformada de Fourier. Uma delas é a Transformada Wavelet Discreta, a qual é capaz de dar uma resolução temporal e uma resolução de frequência para um sinal discreto. Nas próximas seções serão apresentadas as ferramentas necessárias para construir a Transformada Wavelet Discreta, começando com a convolução.



### 2.1.1 Convolução em $\mathbb{R}^N$ .

A operação de convolução de vetores é a base dos filtros lineares. A forma como serão filtrados os dados neste trabalho é por meio de convoluções entre os vetores que representam os sinais e vetores que representam os filtros. O resultado dessa operação é dado pelo Teorema da Convolução, que garante que ao filtrar um vetor, isolamos e modificamos a amplitude das frequências que o compõe, de acordo com alguma necessidade previamente estabelecida.

Suponha que  $x$  é um vetor de dimensão infinita <sup>5</sup>, limitado, com ou sem norma  $\ell_2$  finita, que é denotado por  $\{x_t\}_{t=-\infty}^{+\infty}$ . Em vez de utilizar os valores de  $x$ , pode-se por meio de médias e diferenças analisar os valores em cada posição juntamente com valores próximos. Por exemplo, um dos instrumentos de análise técnica utilizado nos mercados financeiros são as médias móveis que consistem em tomar médias de diferentes tamanhos, i.e. a média das últimas 10, 20, 30... observações do preço de um ativo, e supor que existe alguma informação útil a ser obtida do vetor das médias, especialmente a informação gerada ao se comparar duas médias de tamanho diferentes. Segundo Gençay, Selçuk e Witcher [49], médias com tamanhos menores captam movimentos de mudança de tendência mais rapidamente do que uma média de tamanho maior. Utilizadas em conjunto elas geram informações sobre oportunidades de compra e venda de um ativo.

Matematicamente escrevem-se estas médias<sup>6</sup> (aritméticas) como

$$ma_x(K, t) = \frac{1}{K} \sum_{h=0}^{K-1} x_{t-h}$$

onde  $K$  é o número de observações usadas na computação da média e  $t$  denota o tempo. Aqui o peso dado a cada observação é igual. É possível também dar pesos e sinais diferentes a cada uma das observações, obtendo também diferenças, denotadas por  $d$  e descontar o tempo, fazendo uma sequência de pesos decrescente em  $h$ , pesos exponenciais [45] ou mesmo pesos fora de sequência [20]. Um operador muito utilizado é o da primeira

---

<sup>5</sup>Até aqui foram utilizados apenas vetores pertencentes a espaços de dimensão finita, mas como será visto essa hipótese traz grandes implicações para os resultados obtidos quando filtra-se um sinal por meio de convoluções. Ao utilizar um sinal infinito estas preocupações desaparecem, portanto para fins de ilustração e descrição, primeiramente serão utilizados sinais infinitos

<sup>6</sup>Não confundir esta *moving average* com modelos econométricos MA

diferença:

$$d(x_t) = x_t - x_{t-1} \quad (2.11a)$$

Note que se for tomada a primeira diferença da primeira diferença,  $d(d(x_t))$ , tem-se

$$d(d(x_t)) = d(x_t) - d(x_{t-1}) = x_t - 2x_{t-1} + x_{t-2}. \quad (2.11b)$$

Em geral chamam-se estes tipos de operadores (médias e diferenças) de operadores de convolução, dos quais a média móvel é um caso particular [55]. Sejam  $x$  e  $w$  dois vetores de dimensão infinita, a convolução entre eles é denotada  $(w * x)$ . Ela deve ser calculada entrada a entrada. Para a entrada na posição 0 tem-se

$$g(0) = x * w = \sum_{h=-\infty}^{+\infty} x_{-h}w_h. \quad (2.12)$$

Uma questão importante é saber quando que a convolução está definida. Para vetores de dimensão infinita, é suficiente que  $x$  e  $w$  pertençam a  $\ell_2^7$ , ou que um dos vetores tenha suporte compacto, isto é, um número finito de entradas diferentes de 0 [9].

Se for utilizada a média móvel,  $w$  representa os pesos (que somam um) dados a cada observação. A equação (2.12) retorna um único valor, no caso a média centrada em  $t = 0$  do vetor  $x$ . Se quisermos obter a média (ou diferença) centrada em  $t'$ , escreveríamos

$$g(t') = x * w(t') = \sum_{h=-\infty}^{+\infty} x_{t'-h}w_h = \sum_{h=-\infty}^{+\infty} x_hw_{t'-h}. \quad (2.13)$$

O conjunto de todos os valores  $\{g(\cdot)\}$  calculados com (2.12) é o vetor resultante da operação de convolução. Um instante de reflexão nos leva a concluir que a operação  $*$  é comutativa. Note que o termo à direita de (2.13) é igual a  $w * x(t')$ , que se multiplicarmos  $x$  ou  $w$  por quaisquer escalares, a convolução também vai ser multiplicada por estes escalares e se  $y$  também for um vetor, então

$$(x + y) * w = x * w + y * w.$$

Assim o operador de convolução é um operador linear para cada  $w$  fixo.

Se for relaxada a hipótese de que o vetor  $x$  é infinito um problema é imediato. Seja  $x$  um vetor de tamanho  $N$  e suponha que o vetor  $w$  tenha um número ímpar de

---

<sup>7</sup>Isto é, tenham entradas cujos quadrados são somáveis;  $\sum_{t=-\infty}^{+\infty} x_t^2 < \infty$ .

entradas  $L$ ,  $L < N$ . O primeiro valor da média móvel que vai ser possível calcular será quando  $t = \frac{L+1}{2}$ , e o último quando  $t = N - \frac{L+1}{2}$ . Isso significa que alguma informação vai ser descartada, o que em séries temporais pequenas pode ser custoso ou quando precisamos de informações atualizadas, existe um *lag*. Isso também limita o tamanho da média móvel que podemos utilizar dependendo dos dados. Na verdade, estritamente falando, a convolução nem está definida se  $x$  não for infinito [9].

Uma solução que geralmente é adotada (referências [9, 58, 49, 46]) é utilizar o que é chamado convolução circular, para um vetor de pesos  $w$  com  $L$  entradas diferentes de zero, dada por

$$x * w(t') = \sum_{h=0}^{L-1} [x_{(t'-h) \bmod N}] w_h. \quad (2.14)$$

A equação (2.14) considera que o vetor  $x$  é periódico de período  $N$ . Além disso ela é o que se chama de convolução causal, que só considera valores contemporâneos ou passados (desde que  $x$  seja periódico) para o cálculo da média móvel. Nota-se que os valores resultantes da convolução circular também são periódicos com período  $N$ , portanto é necessário calcular apenas este número de valores (i.e. fazer  $t'$  variar de 0 a  $N-1$ ) pois a convolução circular de tamanho finito resulta nos mesmos valores que uma convolução em um vetor infinito periódico de período  $N$ .

A aplicação da convolução circular forma um vetor em  $\mathbb{R}^N$ , que será denotado por  $\{g\}_{k=0}^{N-1}$ . Pode-se escrever  $g$  em notação matricial como um sistema  $g = Wx$ , ou

$$g = \begin{bmatrix} w_0 & w_{N-1} & w_{N-2} & \dots & w_2 & w_1 \\ w_1 & w_0 & w_{N-1} & \dots & w_3 & w_2 \\ \vdots & & \ddots & & & \vdots \\ w_{N-1} & w_{N-2} & w_{N-3} & \dots & w_1 & w_0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix} \quad (2.15)$$

onde as linhas de  $W$  são translações *mod N* para a direita da primeira linha, obedecendo a regra do módulo  $N$ . A equação (2.15) será denotada por  $(x * w)$ , que a partir de agora vai sempre significar convolução circular. Caso  $L < N$  então  $w_k = 0$  se  $k > L-1$ . Estendendo (2.14) periodicamente indefinidamente, se vê que a convolução circular também é um operador linear para cada  $w$  fixo, pelos mesmos motivos da convolução em dimensão infinita.

**Definição 2.** O vetor  $g \in \mathbb{R}^N$  definido em (2.15) é dito **convolução circular** de  $x$  por  $w$ .

Cabe salientar que supor periodicidade é uma maneira de completar a série temporal para poder calcular a convolução, já que em geral estão disponíveis apenas um número limitado de dados. Existem outras maneiras de completá-la que são bastante úteis e podem ser melhor adequadas ao problema que se quer resolver [9, 49]. Quando forem utilizados dados reais será comentado uma outra forma de se completar a série.

Com o que foi visto até aqui pode-se provar o que talvez é o teorema mais importante da teoria de processamento de sinais, o Teorema da Convolução. A prova que segue foi dada por B. R. Hunt [32] em 1971.

**Teorema 1** (Teorema da Convolução, por B. R. Hunt [32]). *Sejam  $\{x\}_0^{N-1}$  e  $\{w\}_0^{N-1}$  dois vetores com entradas reais, de transformadas discretas de Fourier  $\{\mathbf{X}_k\}_{k=0}^{N-1}$  e  $\{\mathbf{W}_k\}_{k=0}^{N-1}$  respectivamente. Então a transformada discreta de Fourier da convolução circular entre  $x$  e  $w$  será, entrada a entrada,  $\mathbf{X}_k \mathbf{W}_k$ .*

*Demonstração.* Uma matriz  $W$  da mesma forma que (2.15), na qual cada diagonal tem sempre as mesmas entradas, é chamada de matriz circulante. A prova do teorema usa o fato de que estas matrizes são diagonalizáveis e seus autovetores são os vetores da base de Fourier. Seja  $W$  a matriz circulante (2.15) e sejam  $c(k)$ ,  $k = 0, 1, \dots, N - 1$  as  $N$  raízes da unidade,

$$c(k)^N = 1.$$

Pela formula de De Moivre<sup>8</sup>

$$c(k) = e^{\frac{2\pi ik}{N}}$$

---

<sup>8</sup> $[\cos(x) + i\sin(x)]^n = \cos(nx) + i\sin(nx)$

para  $k = 0, 1, 2, \dots, N - 1$ . Para cada  $k$ , os  $N$  possíveis valores de  $c(k)$  formam o vetor  $E_k$  da Base de Fourier. Define-se para um vetor  $p(t)$ ,  $t = 0, 1, \dots, N - 1$ ,

$$\lambda(k) = p(0) + p(1)c(k) + p(2)c(k)^2 + p(3)c(k)^3 + \dots + p(N - 1)c(k)^{N-1} \quad (2.16)$$

que satisfaz o seguinte sistema linear

$$\begin{aligned} \lambda(k) &= p(0) + p(1)c(k) + p(2)c(k)^2 + \dots + p(N - 1)c(k)^{N-1} \\ \lambda(k)c(k) &= p(0)c(k) + p(1)c(k)^2 + p(2)c(k)^3 + \dots + p(N - 2)c(k)^{N-1} + p(N - 1) \\ &\vdots \\ \lambda(k)c(k)^{N-1} &= p(0)c(k)^{N-1} + p(1)c(k)^{N-2} + p(2)c(k)^{N-3} + \dots + p(N - 2)c(k) + p(N - 1) \end{aligned}$$

escrito em formato matricial,

$$P \begin{bmatrix} 1 \\ c(k) \\ c(k)^2 \\ \vdots \\ c(k)^{N-1} \end{bmatrix} = \lambda(k) \begin{bmatrix} 1 \\ c(k) \\ c(k)^2 \\ \vdots \\ c(k)^{N-1} \end{bmatrix} .$$

Vemos que  $P$  é uma matriz circulante cujas entradas são rotações do vetor  $[p(0), p(1), \dots, p(N - 1)]$  e ela possui  $N$  autovetores dados pelos  $E_k$ , cada um associado ao autovalor  $\lambda(k)$ . A matriz

$$E = [E_0, E_1, E_2, \dots, E_{N-1}]$$

possui colunas ortogonais entre si e temos pelo teorema da diagonalização de matrizes [42]

$$P = EDE^{-1}, \quad (2.17)$$

onde  $D$  é uma matriz diagonal cuja  $k$ -ésima entrada é  $\lambda(k)$ . Aplicando esse fato a (2.15), com  $P = W$ :

$$g = EDE^{-1}x \quad (2.18a)$$

$$E^{-1}g = DE^{-1}x. \quad (2.18b)$$

Já que a matriz  $E$  é ortogonal com entradas complexas, sua inversa é obtida multiplicando-se sua hermitiana <sup>9</sup> por  $\frac{1}{N}$ . O lado esquerdo de (2.18b) fica, entrada a entrada,

$$[E^{-1}g]_k = \frac{1}{N} \sum_{t=0}^{N-1} g_t e^{\frac{-2\pi ikt}{N}}. \quad (2.19)$$

O lado direito de (2.19) é a transformada de Fourier de  $g$  multiplicada por  $\frac{1}{N}$ . Semelhantemente temos  $[E^{-1}x]_k = \sum_{t=0}^{N-1} x_t e^{\frac{-2\pi ikt}{N}}$  no lado direito de (2.19). Falta saber o que acontece com a matriz  $D$ . A  $k$ -ésima entrada da diagonal principal é dada pela equação (2.16), onde as potências de  $c(k)$  são as entradas de  $E_k$ , e os  $p(j)$  são as entradas da primeira linha de  $P$ :

$$\lambda(k) = \frac{1}{N} \sum_{t=0}^{N-1} [P_{1t}] e^{\frac{2\pi ikt}{N}}.$$

Substituindo pelos valores de  $W$  nota-se que a sua primeira linha tem entradas dadas por  $w_{-t \bmod N}$ . Com isso a equação fica

$$\lambda(k) = \frac{1}{N} \sum_{t=0}^{N-1} w_{-t \bmod N} e^{\frac{2\pi ikt}{N}},$$

que é igual a

$$\lambda(k) = \frac{1}{N} \sum_{t=0}^{N-1} w_t e^{\frac{-2\pi ikt}{N}}. \quad (2.20)$$

Com isso mostra-se que as entradas da diagonal da matriz  $D$  são as transformadas de Fourier do vetor  $w$  e prova-se o teorema. Especificamente, dada uma matriz de convolução, a transformada de Fourier do vetor  $Wx$  é a transformada de Fourier da primeira linha de  $W$  vezes a transformada de Fourier de  $x$ .

□

A partir desse momento passamos a nos referir ao vetor  $w$  como um filtro, e chamamos a operação de convolução de  $x$  por  $w$  de filtragem. A razão é simples: se for conhecida sua DFT então para qualquer vetor  $x$  que for convolvido com  $w$  teremos um

---

<sup>9</sup>A transposta de  $E$  conjugada.

vetor cujo conteúdo de frequência é uma ponderação das frequências que formam  $x$ . Para ver o porquê, define-se a função ganho de um filtro [46]. Lembrando que qualquer número complexo  $a + bi$  pode ser escrito como  $\sqrt{a^2 + b^2}e^{i\theta}$  (onde  $\theta$  é o arco, com sentido dado pelo sinal de  $b$ , cuja tangente é  $\frac{b}{a}$ , e tomando-se o limite quando  $a \rightarrow 0^+$  ou  $a \rightarrow 0^-$ ) escrevemos  $\mathbf{W}_k$  como uma função de  $k$  da seguinte forma:

$$\mathbf{W}_k = a(k) + ib(k) = |A(k)|e^{i\theta(k)}.$$

A função  $A(k)$  é chamada **função ganho** e  $e^{i\theta(k)}$  é chamada de **função fase** do filtro. Se  $g = x * w$ , o Teorema da Convolução nos diz que sua DFT será entrada a entrada

$$\mathbf{G}_k = |A(k)|e^{i\theta(k)}\mathbf{X}_k. \quad (2.21)$$

Pelo Teorema de Parseval,

$$\|g\|^2 = \frac{1}{N}\|\mathbf{G}\|^2 = \frac{1}{N}\sum_{k=0}^{N-1}|A(k)|e^{i\theta(k)}\mathbf{X}_k\overline{|A(k)|e^{i\theta(k)}\mathbf{X}_k} = \frac{1}{N}\sum_{k=0}^{N-1}|A(k)|^2|\mathbf{X}_k|^2.$$

Disso deduzimos que se  $|A(k)|$  for pequeno o suficiente o vetor  $g$  vai ter pouca influência do componente  $E_k$  presente em  $x$ . Reciprocamente, quanto maior  $|A(k)|$  mais influência terá a frequência  $k$  no vetor filtrado. A função ganho quadrado será denotada pelo símbolo estilizado de sua DFT, seguindo a notação de Percival e Walden [46], e.g.  $\mathcal{A}(k)$ .

**Definição 3.** A função ganho de um filtro é o quanto ele altera a amplitude das frequências que compõem um sinal. Uma aplicação do Teorema de Parseval e do teorema da convolução dizem que se  $g$  é o vetor resultante da convolução entre  $x$  e um filtro com função ganho  $A(k)$ , então  $\|g\|^2 = \frac{1}{N}\sum_{k=0}^{N-1}|A(k)|^2|\mathbf{X}_k|^2 = \frac{1}{N}\sum_{k=0}^{N-1}\mathcal{A}(k)|\mathbf{X}_k|^2$ .

Quanto à função fase, a inversão da transformada de Fourier de  $g$  implica que:

$$g = \frac{1}{N}\sum_{k=0}^{N-1}\mathbf{G}_k E_k.$$

Agora, uma forma de se interpretar a transformada de Fourier de um vetor é como sendo uma combinação de funções sinusoides, onde os coeficientes representam a amplitude

e o deslocamento (atraso) com os quais cada onda aparece. Por *deslocamento* entende-se expressões do tipo  $\cos(t + \theta)$ , onde  $\theta$  é um ângulo medido em radianos. Essa função tem o mesmo período de  $\cos(t)$  mas seu gráfico aparece deslocado para a direita. Em se tratando de vetores, isso equivale a uma rotação, i.e.  $(x_0, x_1, x_3, \dots) \rightarrow (x_{N-1}, x_0, x_1, \dots)$  que é um atraso (no tempo) dos valores do sinal.

Analisando (2.21), especificamente o termo  $e^{i\theta}$  (suprimindo a relação com  $k$ ), podemos reescrever a expressão de  $\mathbf{G}_k$  como o produto de uma função que depende da frequência  $k$  por  $e^{i\theta}$ . Explicitamente:

$$\mathbf{G}_k E_k = A(k) e^{i\theta} E_k. \quad (2.22)$$

O vetor  $e^{i\theta} E_k$  é dado, entrada a entrada, por  $e^{i\theta} e^{\frac{2\pi i k t}{N}}$  e substituindo  $\frac{2\pi i k t}{N}$  por  $\alpha$ , pela fórmula de Euler e pelas relações trigonométricas, temos:

$$\begin{aligned} [e^{i\theta} E_k]_t &= [\cos(\theta) + i\text{sen}(\theta)][\cos(\alpha) + i\text{sen}(\alpha)] \\ &= \cos(\theta)\cos(\alpha) + i\cos(\theta)\text{sen}(\alpha) + i\cos(\alpha)\text{sen}(\theta) - \text{sen}(\theta)\text{sen}(\alpha) \\ &= \cos(\alpha + \theta) + i\text{sen}(\alpha + \theta), \end{aligned}$$

que é simplesmente uma versão transladada<sup>10</sup> de  $E_k$ . Com isso, as equações (2.21) e (2.22) implicam que um filtro ao ser aplicado a um vetor, pode ser descrito como uma alteração das amplitudes das frequências junto com um atraso (ou avanço se  $\theta$  for negativo).

**Definição 4.** A função fase de um filtro é o quanto ele atrasa ou avança no tempo um sinal filtrado. Sem perda de generalidade, suponha que a DFT de  $x$  seja constante e igual a 1 para todo  $k$ , isto é  $x$  é a soma dos  $E_k$ 's e que  $w$  é um filtro com DFT dada, para cada  $k$ , por  $\mathbf{W}_k = |A(k)|e^{i\theta}$ . Então  $g = x * w$  tem DFT dada por  $|A(k)|e^{i\theta}$ . A multiplicação de cada  $E_k$  por  $e^{i\theta}$  na decomposição de  $g$  na base de Fourier equivale a um atraso ou avanço (dependendo do sinal de  $\theta$ ) nos valores de suas entradas.

<sup>10</sup>Tecnicamente para ser uma translação,  $\theta$  deveria ser um múltiplo de  $\frac{\pi}{N}$  mas de qualquer forma os valores de  $x$  vão sofrer um atraso junto com a mudança das amplitudes de seus componentes de frequência.



Para dar um exemplo concreto, um filtro particularmente importante é o filtro de Haar, também chamado wavelet de Haar. Este filtro tem apenas duas entradas diferentes de zero,  $w_0 = \frac{1}{\sqrt{2}}$  e  $w_1 = \frac{-1}{\sqrt{2}}$ , o que faz dele um dos filtros mais simples que existem <sup>11</sup>. Ele também é bastante intuitivo, sendo simplesmente a diferença aritmética, multiplicada por uma constante, entre dois valores de  $x$ . Note que enquanto  $w$  tem apenas duas entradas, queremos filtrar um vetor de tamanho arbitrário  $N$ . Para fazer isso, simplesmente completamos o vetor  $w$  com tantos zeros quanto forem necessários de forma a obter um vetor em  $\mathbb{R}^N$  com o qual forma-se a matriz em (2.15). No caso do **filtro de Haar**, tem-se:

$$(w_0, w_1) \longrightarrow (w_0, \underbrace{0, 0, 0, \dots, 0}_{N-2 \text{ zeros}}, w_1) \quad (2.23)$$

O vetor em (2.23) está apto a ser utilizado em uma convolução circular. As entradas de  $x * w$  serão dadas por  $\frac{x_t - x_{t-1 \bmod N}}{\sqrt{2}}$  para  $t = 0, 1, \dots, N - 1$ . Este procedimento é adotado em todos os filtros que serão utilizados mais tarde. Para calcular a função ganho do filtro define-se  $\phi(k) = \frac{2\pi i k(N-1)}{N}$  e como  $w$  só tem a primeira e a última entradas diferentes de zero sua DFT é dada por

$$\mathbf{W}_k = \frac{1}{\sqrt{2}} - \frac{\cos(\phi) - i \operatorname{sen}(\phi)}{\sqrt{2}}.$$

Tomando o produto de  $\mathbf{W}_k$  pelo seu conjugado, chega-se a expressão

$$\mathcal{W}(k) = 1 - \cos(\phi).$$

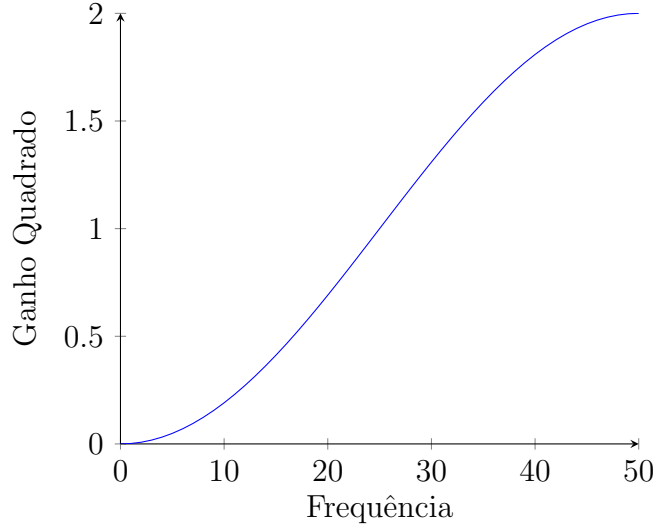
A Figura 2.4 apresenta o gráfico da função  $\mathcal{W}(k)$  para  $N = 100$ , com frequência de Nyquist igual a 50.

Como  $\cos(\phi)$  assume uma vez cada valor entre  $-1$  e  $1$  no intervalo  $[0, \pi]$ , independente do valor de  $N$ , a função  $\mathcal{W}$  vai manter o mesmo formato, não importando o tamanho do sinal. Isso é conveniente por que dessa forma sabemos que estamos sempre filtrado um sinal da mesma forma relativa em relação a frequência de Nyquist. Frequências que estão no intervalo entre  $\alpha f_N$  e  $f_N$ , para  $\alpha$  entre 0 e 1, serão sempre filtradas com o mesmo ganho. Nota-se que uma frequência  $k$  qualquer pode estar em intervalos diferentes de acordo como tamanho do sinal, já que a suposição do Teorema de Shannon-Nyquist diz respeito somente à taxa de amostragem  $N$ , não à quantidade de amostras  $k$ .

---

<sup>11</sup>É possível ver a identidade como um filtro que mantém as amplitudes e fases iguais, e pela linearidade da convolução, uma constante vezes a identidade modifica a amplitude de todas as frequências igualmente.

**Figura 2.4:** Função ganho quadrado da Wavelet de Haar



Fonte: elaborado pelo autor.

Uma propriedade da função ganho que pode ser observada no gráfico de  $\mathcal{W}$  é que a função é estritamente crescente no intervalo entre zero e  $f_N$ . Além disso quando  $k = 0$ , tem-se  $W_0 = 0$ , ou seja, a wavelet de Haar zera a a frequência constante e amplifica a frequência de Nyquist:  $\mathbf{W}_0 \mathbf{X}_0 = 0$  e  $\mathbf{W}_{N-1} \mathbf{X}_k = 2e^{i\theta} \mathbf{X}_k$  para qualquer  $x$ . Um filtro que tem esta última propriedade é chamado de **passa-alto** (pa), ele é um filtro que deixa passar apenas as frequências mais altas [9]. Da mesma forma, se um filtro zera a frequência de Nyquist e mantém a constante, ele é chamado de **passa-baixo** (pb). Broughton e Bryan [9] utilizam o que se chama Transformada Z para formalizar o que se requer de um filtro para ele ser um filtro pa ou pb. É possível traduzir isso para a álgebra matricial de forma simples, bastando notar que os vetores  $E_0$  e  $E_{\frac{N}{2}}$  são dados, posição a posição, por

$$[E_0]_j = 1$$

e

$$[E_{\frac{N}{2}}]_j = (-1)^j,$$

não importando qual  $N$ . Portanto para ser um pb, o requerimento é que o coeficiente de Fourier devido à frequência  $k = N/2$  dado por  $\langle w, E_{\frac{N}{2}} \rangle = w_0 - w_{N-1} + w_{N-2} - \dots - w_1$  seja igual a 0. Já o coeficiente de Fourier devido à frequência  $k = 0$ , dado por  $\langle w, E_0 \rangle = w_0 + w_{N-1} + w_{N-2} + \dots + w_1$ , deve ser diferente de zero. Essas são as mes-

mas equações dadas por Broughton e Bryan [9, p. 243] utilizando a transformada Z. Elas podem ser sumarizadas (para um filtro pa) como

$$\sum_{j=0}^{N-1} w_j = 0 \quad (2.24a)$$

e

$$\sum_{j=0}^{N-1} w_j (-1)^j \neq 0. \quad (2.24b)$$

Para um filtro ser pb é só trocar as igualdades e desigualdades acima entre si. Essa é uma regra fácil de se verificar para saber quando um filtro é pb ou pa.

Existe um filtro pb de Haar que é usado em conjunto com a wavelet de Haar, chamado função escala de Haar <sup>12</sup>, novamente com duas entradas,  $g_0 = \frac{1}{\sqrt{2}}$  e  $g_1 = \frac{1}{\sqrt{2}}$ . Nesse caso, o filtro são os valores médios (com peso somando  $\sqrt{2}$ ) entre dois valores subsequentes de  $x$ . Os filtros passa alto serão denotados por  $g$ .

Substituindo a diferença pela soma no cálculo da DFT da wavelet de Haar, conclui-se que  $\mathcal{G}(k)$ , a função ganho quadrado de  $g$  é dada por

$$\mathcal{G}(k) = 1 + \cos(\phi).$$

A Figura 2.5 apresenta o gráfico de  $\mathcal{G}$ . Note que ele é similar ao gráfico de  $\mathcal{W}$  no sentido de que um é uma reflexão deste. Isto é,  $\mathcal{G}$  é uma reflexão de  $\mathcal{W}$  a partir do ponto  $f_N$ . Isso é um propriedade da wavelet e da função escala de Haar: a soma de suas funções ganho são sempre iguais a  $\sqrt{2}$ , ou no caso do ganho quadrado, iguais a 2. Essa relação entre os dois filtros é chamada de quadratura, e eles recebem o nome de filtros espelhos de quadratura, que pode ser expressa para quaisquer filtros  $w$  e  $g$  tais que  $\mathcal{W}(k) + \mathcal{G}(k) = C$  [46], onde  $C$  é uma constante arbitrária que depende dos filtros mas não de  $k$ . Com uma reescala dos filtros (dividindo por uma constante apropriada) é sempre possível escolher qualquer valor desejado para  $C$ , já que a DFT é linear.

---

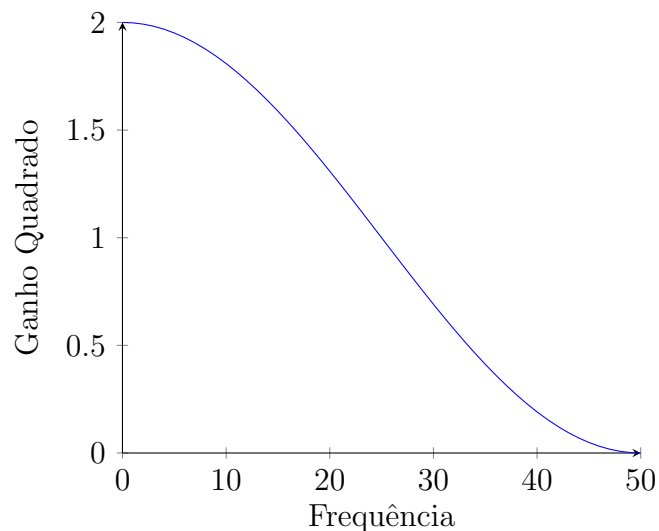
<sup>12</sup>Essa nomenclatura é usada devido ao trabalho seminal de Haar [30]

**Definição 5.** Um filtro que atenua as frequências mais altas que compõe um vetor e que amplifica as frequências mais baixas é chamado de filtro passa-baixo.

**Definição 6.** Um filtro que atenua as frequências mais baixas que compõe um vetor e que amplifica as frequências mais altas é chamado de filtro passa-alto.

**Definição 7.** Dois filtros que ao serem utilizados em conjunto amplificam ou atenuam igualmente todas as frequências que compõem um vetor são chamados filtros espelho de quadratura

**Figura 2.5:** Função ganho quadrado da Função Escala de Haar



Fonte: elaborado pelo autor.

Quanto à fase, ela é mais complicada, envolvendo a fatoração do valor da DFT para encontrar uma função que dependa do produto de uma exponencial em  $\theta$  e da função ganho. Felizmente existe uma forma que será utilizada mais tarde de não ter de se preocupar com a fase de um filtro, chamada de correlação cruzada [46].

Para dar um exemplo prático de filtragem, nos baseamos em Broughton e Bryan [9, p. 203]. Suponha que o sinal  $x$  seja dado por

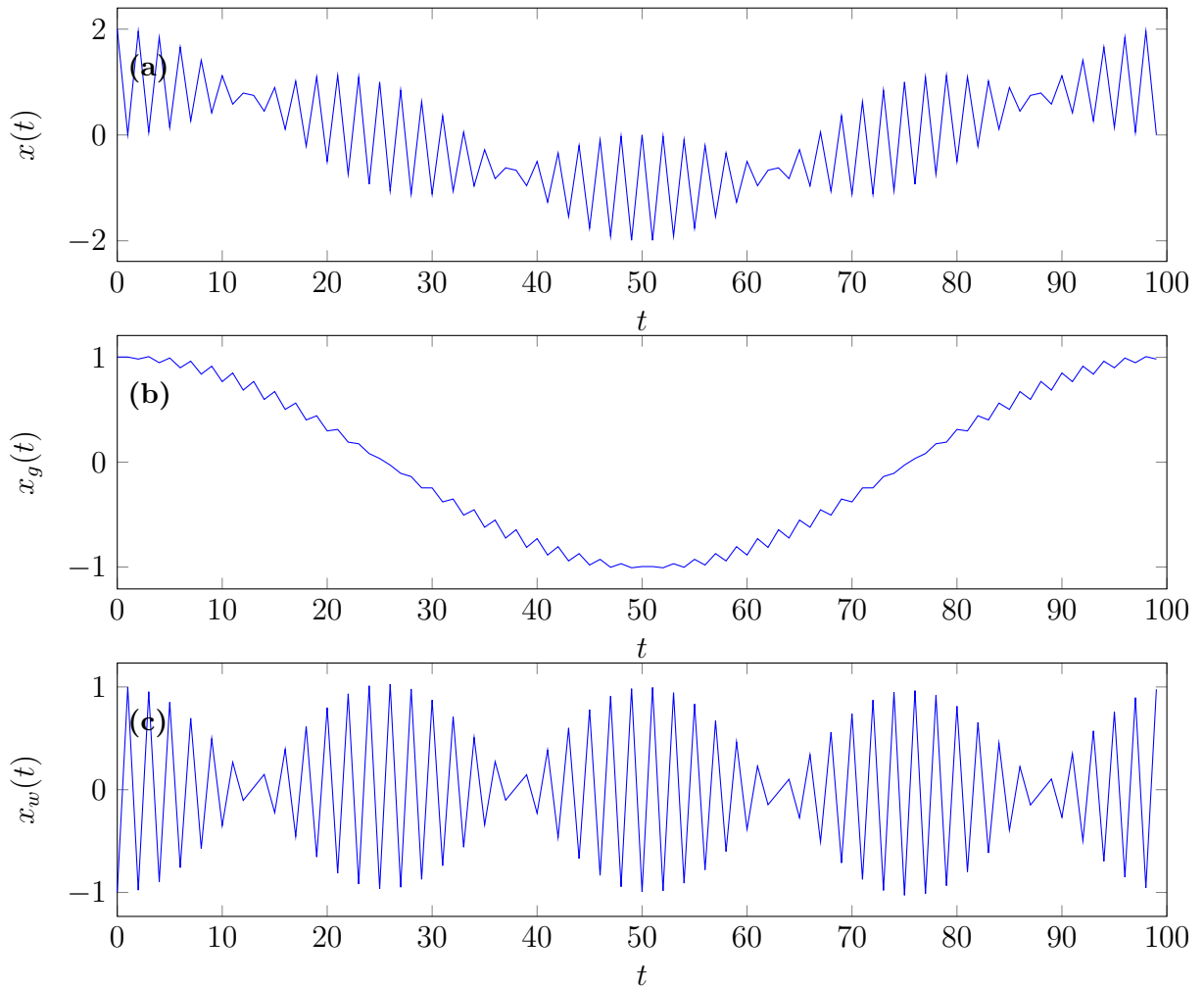
$$x(t) = \cos\left(\frac{2\pi t}{N}\right) + \cos\left(\frac{96\pi t}{N}\right) \quad (2.25)$$

com  $N = 100$  e  $t$  variando de zero a 99. Isto é, amostramos um sinal composto por duas funções cosseno a uma taxa de 100 observações no intervalo  $[0, 1]$ . Redimensionamos os coeficientes dos filtros para  $w = \left(\frac{1}{2}, \frac{-1}{2}\right)$   $g = \left(\frac{1}{2}, \frac{1}{2}\right)$ , de acordo com a proposta de Percival e Walden [46]<sup>13</sup>, de forma que  $\mathcal{W} + \mathcal{G} = 1$ . Dessa forma, os vetores  $x_\ell$  e  $x_h$  representam  $x$  filtrado por  $g$  e  $w$ , respectivamente. A figura 2.6 apresenta os três vetores, o original (2.6a), o filtrado por  $g$ , (2.6b), e o filtrado por  $w$ , (2.6c). Como se pode ver, os filtros isolam cada componente de  $x$  (96 é duas vezes a frequência 48, logo é isolado por  $w$ ), mas de forma imperfeita. A forma serrilhada da figura 2.6b é devida ao fato de que  $g$  não zera frequência  $k = 48$ , apenas atenua ela. Observando o gráfico 2.5 conclui-se que quanto mais perto de  $k = 25$  uma frequência, menos eficiente vai ser este filtro em atenuá-la.

---

<sup>13</sup>A razão dessa reescala será abordada mais tarde

**Figura 2.6:** Sinal  $x(t)$  filtrado.



**Fonte:** Elaborado pelo autor, baseado em Broughton e Bryan [9]

O filtro de Haar decompõe (aproximadamente) um sinal em bandas, cada uma devida a uma parte de seu espectro. Essa decomposição também é chamada de banco de filtros e as bandas são chamadas de canais [9]. No caso de Haar cada canal representa aproximadamente metade das frequências que compõe o sinal. A função escala de Haar filtra (aproximadamente) as frequências no intervalo  $[0, 1/4]$  e a wavelet de Haar filtra (aproximadamente) as frequências no intervalo  $[1/4, 1/2]$ . Pelo resto deste trabalho, todos os filtros dividirão aproximadamente o espectro do sinal em bandas desse formato. Os filtros  $pb$  filtrarão a primeira metade do espectro e os filtros  $pa$  filtrarão a segunda. Todavia, é possível utilizar filtros que dividem o espectro de outras formas, por exemplo o filtro de Hodrick-Prescott [31].

### 2.1.2 Correlação Cruzada

Uma operação semelhante à convolução é dada pela correlação cruzada [46]<sup>14</sup> entre  $x$  e  $w$ , denotada por  $(x \star w)$ , e definida como

$$(x \star w)(t) = \sum_{k=0}^{N-1} x_{t+k} w_k. \quad (2.26)$$

Uma maneira de interpretar a correlação cruzada é notar que se o vetor  $w$  é escrito como  $(w_0, w_{N-1}, \dots, w_1)$  então a correlação cruzada entre  $x$  e  $w$  é a convolução entre  $x$  e o vetor  $w' = (w_0, w_1, \dots, w_{N-1})$ . Seguindo a sugestão de Percival e Walden [46], a transformada de Fourier de  $w'$  é, entrada a entrada, dada por

$$\mathbf{W}'_k = \sum_{t=0}^{N-1} w_t e^{-\frac{2\pi i k t}{N}},$$

multiplicando o lado direito da equação acima por  $e^{\frac{2\pi i k N}{N}}$ , que é igual a 1 para qualquer  $k$  inteiro, temos:

$$\sum_{t=0}^{N-1} w_t e^{-\frac{2\pi i k t}{N}} = \left( \sum_{t=0}^{N-1} w_t e^{-\frac{2\pi i k t}{N}} \right) e^{\frac{2\pi i k N}{N}}.$$

Chegamos à conclusão que

$$\mathbf{W}'_k = \sum_{t=0}^{N-1} w_t e^{\frac{2\pi i k (N-t)}{N}}. \quad (2.27)$$

Comparando com a transformada de Fourier de  $w$  (equação (2.4)) conclui-se que  $W'_k = \overline{W_k}$ . Em particular, a seguinte observação vai ser utilizada mais adiant.: se for realizada a convolução entre  $w'$  e  $(x \star w)$  então com duas aplicações do Teorema da Convolução, a DFT  $\mathbf{G}'$  do vetor  $g'$  resultante será dada, entrada a entrada, por:

$$\mathbf{G}'_k = \overline{\mathbf{W}_k} \mathbf{W}_k \mathbf{X}_k = |\mathbf{W}_k|^2 \mathbf{X}_k \quad (2.28)$$

Portanto a aplicação de um filtro formado pela auto correlação cruzada de um filtro causal por ele mesmo resulta em um filtro com fase zero. A equação (2.28) vai ter um papel importante na aplicação de filtros às séries temporais por que permite, até certo ponto, ignorar problemas de fase criados pelos filtros.

---

<sup>14</sup>Em [46], os autores utilizam a correlação cruzada complexa dada por  $(x \star \bar{w})$ , mas como serão utilizados apenas filtros com valores reais, não é necessário se preocupar com conjugados e valores complexos, além das funções sinusoides.

### 3 BANCOS DE FILTROS E A TRANSFORMADA WAVELET

Quando foi apresentada a transformada discreta de Fourier de um sinal, é implícita a hipótese de que ele é periódico no tempo, uma hipótese que pode ser bastante irreal. Além disso a transformada de Fourier não tem nenhuma resolução temporal, como visto no exemplo de um sinal formado por diferentes funções em cada metade de seu tempo de duração. Mesmo usando filtros lineares, em nenhum momento foi mostrado que é possível escapar dessa dificuldade. Tudo que o Teorema da Convolução diz é que é possível modificar o conteúdo das frequências. Por isso uma forma diferente de tratar com sinais é proposta pela transformada wavelet.

Continua-se supondo que um sinal é amostrado a uma taxa de  $N$  observações ao longo de um intervalo de tempo dado, normalizado para  $[0, 1]$ . Se for utilizada a linguagem da álgebra linear (ver Lax [41]) os valores do sinal  $x = (x_0, x_1, \dots, x_{N-1})$  observados são obtidos expressos na base canônica de  $\mathbb{R}^N$  e ao expressar o vetor na forma  $x = \frac{1}{N} \sum_{i=0}^{N-1} \mathbf{X}_i E_i$  ele está sendo reescrito em outra base do espaço. Se for desejado, podemos executar a DFT em um vetor, obter os coeficientes, modificá-los e reconstruir um vetor com estes coeficientes modificados, que por (2.10) pode ser mais ou menos próximo de  $x$  de acordo com a aplicação ou com a precisão desejada. O Teorema da Convolução garante que um filtro aplicado ao sinal vai realizar o mesmo procedimento, escolhendo automaticamente quais frequências modificar em um sinal. Por outro lado, a base de Fourier não é a única base que possibilita esse procedimento. Uma família de bases particularmente importante são as bases ortogonais que dão origem às transformadas ortogonais.

A base de Fourier, após uma divisão por  $\sqrt{N}$ , [46] pode facilmente ser transformada em uma base ortonormal. Além disso, pelo Teorema de Gram-Schmidt [41, p. 81] dada qualquer base de um espaço com dimensão finita, podemos encontrar uma base ortogonal a partir dela. A vantagem dessas bases é que é fácil a obtenção dos coeficientes de um vetor escrito nelas e que uma transformação linear cuja representação matricial tem as colunas iguais aos vetores da base preserva a norma de um vetor, se nele ela for aplicada. Se uma base é composta por  $[e_0, e_1, \dots, e_{N-1}]$  vetores coluna, com entradas reais



e ortogonais entre sí, e  $A = [e_0 \ e_1 \ \dots \ e_{N-1}]^*$  então

$$A^*A = I_N \tag{3.1a}$$

e

$$\langle Ax, Ax \rangle = \langle x, x \rangle \tag{3.1b}$$

$$\tag{3.1c}$$

A matriz  $A$  é um operador unitário [41] e  $A^*$  é sua hermitiana. Em particular se  $A$  só tiver entradas reais,  $A^* = A^T$ . Uma consequência do uso da equação (3.1b) é que operadores unitários preservam a variância amostral de um vetor. A variância amostral é dada por  $1/N \langle x - \bar{x}, x - \bar{x} \rangle$ , onde  $\bar{x}$  é um vetor cujas as entradas são o valor médio das entradas de  $x$ . Suponha que  $A$  tenha entradas reais, por (3.1b),  $Ax$  tem a mesma variância que  $x$ .

Se  $x = \alpha_0 e_0 + \alpha_1 e_1 + \dots + \alpha_{N-1} e_{N-1}$  então cada  $\alpha$  será dado por

$$\alpha_j = \langle x, e_j \rangle = \langle Ax, Ae_j \rangle = \langle Ax, \mathbf{1}_j \rangle.$$

Onde  $\mathbf{1}_j$  é o vetor da base canônica com apenas a  $j$ -ésima entrada diferente de zero. Assim vetor  $y = Ax$  é tem a  $j$ -ésima entrada coincidindo com os coeficientes de  $x$  expresso na base formada pelas colunas de  $A$ . Tomando o produto de  $A^T$ , por  $y$  tem-se:

$$A^T y = A^T Ax = x. \tag{3.2}$$

O termo central de (3.2) é de extrema importância. Pelo Teorema de Parseval, temos:

$$\|x\|_2^2 = \sum_{j=0}^{N-1} |\alpha_j|^2.$$

Dado um  $\varepsilon$  escolhe-se uma coleção  $J$  dos  $\alpha$ 's tal que  $\delta = \sum_{j=0}^{N-1} |\alpha_j|^2 - \sum_{j \in J} |\alpha_j|^2 \leq \varepsilon$ . Pelo Teorema da Projeção [46, p.44],  $\delta$  é a menor distância possível entre  $x$  e o espaço gerado pelos vetores  $\{e_j\}_{j \in J}$ . Cabe lembrar que como se tratam de vetores em  $\mathbb{R}^N$  sempre existe um  $\varepsilon$  tal que são necessários todos os vetores da base para que a distância entre eles fique tão pequena ou menor do que  $\varepsilon$ .

De interesse para o processamento de sinais é a representação esparsa que alguns vetores possuem em certas bases [44]. *Sparcity* (ou, permitindo-se uma tradução informal, esparsidade) significa a capacidade de se utilizar poucos vetores da base para reconstruir, aceitando alguma perda, o vetor original  $x$ , i.e. a norma (energia) de um vetor estará concentrada em poucos coeficientes. Dependendo do sinal em questão, a base de Fourier pode ser razoavelmente esparsa, por exemplo o vetor no qual foi usado o filtro de Haar da seção anterior, sua DFT têm apenas dois valores, a menos de uma conjugação (já que os coeficientes de Fourier sempre vêm em pares conjugados para sinais reais [9]). No caso geral, um uso da equação (2.10) garante que se as amplitudes de certas frequências forem relativamente grandes em relação às outras, utilizando poucos coeficientes recupera-se  $x$  com um nível de perda potencialmente pequeno. Por outro lado, se o vetor  $x$  tiver saltos muito grandes entre duas coordenadas sucessivas, isso não é mais verdade. Manter apenas algumas frequências do espectro não permite obter um vetor próximo do original. Para processar um sinal, a estratégia adotada é encontrar uma outra base de  $\mathbb{R}^N$  que consiga captar essas descontinuidades e ser razoavelmente esparsa. Aqui, captar as descontinuidades significa ter resolução temporal, ou seja, no momento que um sinal sofre uma mudança estrutural, alguma coisa acontece nos coeficientes da transformada.

Uma classe que tem essas propriedades são as Transformadas Wavelet. Já foi falado em filtro wavelet (de Haar) anteriormente, e de fato existe uma ligação entre os filtro wavelet e a transformada wavelet e é possível apresentar a transformada wavelet ao mesmo tempo como uma transformada linear e um filtro (e.g. Percival e Walden [46]), ou começar como uma transformada qualquer em  $\mathbb{R}^N$ , e falar em frequências mais tarde (como em Walker [58]). Por motivos de facilidade de exposição, será adotada a segunda abordagem.

### 3.1 Wavelets como Transformadas Lineares

Esta e as próximas duas seções são uma amálgama de principalmente três referências: Broughton e Bryan [9], Percival e Walden [46] e Walker[58]. Cada autor apresenta a matéria de um jeito um pouco diferente dos demais, destacando pontos diferentes,

e aqui, resumidamente, será feita uma síntese entre elas. A divisão entre wavelets como filtros e como transformadas lineares é feita por Percival e Walden [46].

A base para o que será visto será o filtro wavelet, um vetor  $w$  com  $L$ , um número par, entradas diferentes de zero com as seguintes propriedades [46]:

$$\sum_{l=0}^{L-1} w_l = 0 \quad (3.3a)$$

$$\sum_{l=0}^{L-1} w_l^2 = 1 \quad (3.3b)$$

$$\sum_{l=0}^{L-1} w_l w_{(l+2n)} = 0, \quad n \in \mathbb{Z}, \quad n \neq 0. \quad (3.3c)$$

Na equação (3.3c) se  $l + 2n > N - 1$  ou se  $l + 2n < 0$ , então  $w_{l+2n} = 0$ . A primeira propriedade é a condição já vista se  $w$  for um filtro passa-alto. As duas outras garantem que a transformada que será criada a partir de  $w$  será ortonormal. A partir de agora será assumido que  $N$  é um múltiplo de uma potência de dois, assim é possível dividi-lo por dois pelo menos uma vez. Na literatura em geral, pede-se primeiramente que  $N$  seja uma potência de dois, mas para esta construção isso não é necessário.

É também assumido que  $L < N$  e cria-se a matriz  $W$  com dimensões  $\frac{N}{2} \times N$  através de rotações ímpares do vetor  $w$  completado por zeros até que seu tamanho seja  $N$ . Denotaremos a  $n$ -ésima linha de uma matriz  $W$  qualquer por  $W_n$ <sup>1</sup>. Assim  $W_n = \mathcal{T}^{2n+1}w$ , onde  $\mathcal{T}$  é a matriz de rotação. Temos:

$$W = \begin{bmatrix} W_0 \\ W_1 \\ \vdots \\ W_{\frac{N}{2}} \end{bmatrix} \quad (3.4)$$

Explicitamente:

$$W = \begin{bmatrix} w_1 & w_0 & w_{N-1} & \dots & w_2 \\ w_3 & w_2 & w_1 & \dots & w_4 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ w_{N-1} & w_{N-2} & w_{N-3} & \dots & w_0 \end{bmatrix}. \quad (3.5)$$

---

<sup>1</sup>O contexto tornará claro em que casos o subscrito indica que estamos tomando a  $n$ -ésima componente de um vetor ou a  $n$ -ésima linha de uma matriz.

Por exemplo, se  $W$  for formada com os coeficientes do filtro de Haar então ela será dada por

$$W = \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & \dots & 0 \\ 0 & 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}_{\frac{N}{2} \times N} \quad (3.6)$$

Aplicando a um vetor  $x$ ,  $N \times 1$ , obtém-se

$$Wx = \begin{bmatrix} \frac{x_1 - x_0}{\sqrt{2}} \\ \frac{x_3 - x_2}{\sqrt{2}} \\ \vdots \\ \frac{x_{N-1} - x_{N-2}}{\sqrt{2}} \end{bmatrix} \quad (3.7)$$

que são as primeiras diferenças em pares do vetor  $x$ . Fica claro nesse caso por que é necessário que  $N$  seja múltiplo de dois.

Neste trabalho serão utilizadas apenas wavelets da família das Daubechies, por serem de fácil implementação e obtenção e terem propriedades estatísticas que serão úteis mais tarde. Além disso wavelets nessa família têm uma interpretação bastante intuitiva. Segundo Percival e Walden [46, p. 60], os valores obtidos por (3.7) são diferenças (com pesos) entre duas médias ponderadas. No caso de Haar, médias pontuais, i.e. os próprios números divididos por  $\sqrt{2}$ . Similarmente, as outras wavelets da família resultam de tomadas sucessivas de diferenças e médias ponderadas entre um número crescente de entradas de  $x$ , de acordo com o suporte de cada wavelet. Seguindo Percival e Walden [46], para  $L = 4$  e  $y_t = ax_t - bx_{t-1}$ , utilizando (2.11b):

$$\begin{aligned} d(d(y_t)) &= y_t + 2y_{t-1} - y_{t-2} \\ &= ax_t - bx_{t-1} + 2(ax_{t-1} - bx_{t-2}) - ax_{t-2} - bx_{t-3} \\ &= ax_t + (2a - b)x_{t-1} + (2b - a)x_{t-2} + bx_{t-3} \\ &= w_0x_t + w_1x_{t-1} + w_2x_{t-2} + w_3x_{t-3} \end{aligned} \quad (3.8)$$

Para qualquer conjunto  $(w_0, w_1, w_2, w_3)$  que satisfaça as condições (3.3) podem-se encontrar  $a$  e  $b$ , pesos para a média ponderada. Esse padrão continua indefinidamente. Se  $L = 6$ , toma-se  $d(d(d(y_t)))$  e encontram-se valores tais que  $y_t = ax_t + bx_{t-1} +$

$cx_{t-2}$  é uma média ponderada que satisfaz algo similar a (3.8). A heurística deste procedimento é tomar médias de valores de um vetor em torno de um ponto e comparar com médias tiradas com valores um pouco afastados, fazendo isso em sequência. Se um sinal sofrer variações bruscas, a tendência é que estas diferenças sejam maiores. Alternativamente, um sinal que sofre poucas variações tende a ter estas diferenças pequenas.

Encontrar os valores de  $w$ , por outro lado, é uma tarefa que possivelmente não pode ser feita com as ferramentas utilizadas até aqui sem confusão notacional. Para uma introdução ao assunto, Broughton e Bryan [9] dedicam uma seção de seu livro para desenhos de filtros, incluindo as Daubechies, com o uso da transformada Z, que simplifica enormemente a notação das operações necessárias e deixa o texto mais legível. Neste trabalho nos contentaremos em descrever as propriedades de cada wavelet utilizada, em momento oportuno.

Como no caso do filtro de Haar, onde a wavelet tem acompanhada de si uma função que chamamos escala, as outras wavelets também possuem uma função escala, que é obtida a partir de seus coeficientes da seguinte forma:

$$g_j = (-1)^{j+1}w_{L-1-j}. \quad (3.9)$$

Para  $j = 0, 1, 2, \dots, L - 1$ . É imediato ver que função escala satisfaz (3.3b) e (3.3c) pois ela tem os mesmos valores de  $w$ , exceto por uma troca de sinal, e mantém a distância relativa entre eles <sup>2</sup>. Ela também satisfaz  $\sum_{l=0}^{L-1} g_l = \sqrt{2}$  [46]. Assim como com  $w$ , é criada uma matriz  $\mathbf{G}$  de tamanho  $\frac{N}{2} \times N$  completando-se  $g$  com zeros até ele ter o tamanho certo e fazendo cada linha de  $\mathbf{G}$  ser uma rotação ímpar do vetor original.

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_0 \\ \mathbf{G}_1 \\ \vdots \\ \mathbf{G}_{\frac{N}{2}} \end{bmatrix} \quad (3.10)$$

---

<sup>2</sup>Isto é, se dois componentes de  $w$  estão a uma distância  $h$  um do outro, então os respectivos componentes em  $g$  também distarão a uma distância  $h$

No caso de Haar:

$$\mathbf{G} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \dots & 0 \\ \vdots & & \ddots & & \vdots & \\ 0 & 0 & 0 & \dots & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}_{\frac{N}{2} \times N}. \quad (3.11)$$

Aplicando ao vetor  $x$  :

$$\mathbf{G}x = \begin{bmatrix} \frac{x_1+x_0}{\sqrt{2}} \\ \frac{x_3+x_2}{\sqrt{2}} \\ \vdots \\ \frac{x_{N-1}+x_{N-2}}{\sqrt{2}} \end{bmatrix}. \quad (3.12)$$

O vetor em (3.12) é uma média dois a dois dos valores de  $x$  com ponderação somando  $\sqrt{2}$ . No caso geral, para qualquer  $L$ , aplicar a matriz de escala significa tomar médias ponderadas dos valores de  $x$  ao redor de um ponto, e por construção, os pesos sempre vão somar  $\sqrt{2}$  [46]. A prova será dada mais adiante.

Os vetores  $\mathbf{W}x$  e  $\mathbf{G}x$  podem ser calculados simultaneamente em uma mesma matriz simplesmente "empilhando"  $\mathbf{G}$  e  $\mathbf{W}$  e criando uma matriz  $\mathbf{P}$   $N \times N$ . Utilizando a notação de Walker [58]:

$$\mathbf{P}x = \begin{bmatrix} \mathbf{G} \\ \mathbf{W} \end{bmatrix} x = \begin{bmatrix} \mathbf{G}x \\ \mathbf{W}x \end{bmatrix} = \begin{bmatrix} \mathbf{a} \\ \mathbf{d} \end{bmatrix}. \quad (3.13)$$

O lado direito da equação (3.13) é representado por  $\mathbf{a} = \mathbf{G}x$ , chamado de vetor de coeficientes *average* e  $\mathbf{d} = \mathbf{W}x$ , chamado de vetor de coeficientes *detail*. Salientando novamente: eles representam a média, ou tendência, e o detalhe (a diferença entre as médias) entre valores de  $x$  ao redor de um ponto. Por exemplo, escrevendo  $\mathbf{a} = (a_0, a_1, \dots, a_{\frac{N}{2}})$ , vemos que cada coordenada  $j$  desse vetor é calculada pelos  $L$  últimos valores de  $x$  a partir de  $x_{2j}$ , e a coordenada seguinte descarta os dois valores mais antigos e adiciona os dois próximos. Isso dá uma ideia de evolução do sinal ao mesmo tempo que podemos localizar cada coeficiente no tempo. Evidentemente o mesmo vale para  $\mathbf{d}$ . A matriz  $\mathbf{P}$  é chamada de **Transformada Wavelet Discreta** (DWT, do inglês *Discrete Wavelet Transform*)

O fato de que a matriz  $\mathbf{P}$  localiza médias e variações em um sinal evidencia o fato de sua utilidade, ademais: esta é uma matriz ortogonal. A prova dada por Percival e

Walden [46] é a seguinte. Como linhas de  $\mathbf{G}$  e  $\mathbf{W}$ , são dadas respectivamente por  $\mathcal{T}^{2n+1}g$  e  $\mathcal{T}^{2n'+1}w$ , onde  $n$  é o número da linha, temos que mostrar que para  $0 \leq t \leq t' \leq \frac{N}{2} - 1$

$$\langle \mathcal{T}^{2t+1}g, \mathcal{T}^{2t'+1}w \rangle = 0.$$

fazendo  $n = t' - t$

$$\langle \mathcal{T}^{2t+1}g, \mathcal{T}^{2t'+1}w \rangle = \langle g, \mathcal{T}^{-2t-1}\mathcal{T}^{2t'+1}w \rangle = \langle g, \mathcal{T}^{2n}w \rangle = \sum_{j=0}^{\frac{N}{2}-1} g_j w_{j+2n}.$$

Utilizando (3.9) Percival e Walden [46, p. 514] demonstram que o lado direito da equação é igual a zero para  $L < N$ .<sup>3</sup>

**Definição 8.** A matriz de rotação  $\mathcal{T}$  troca de ordem as coordenadas de um vetor de forma que a última coordenada vira a primeira, a primeira vira a segunda, a segunda a terceira e assim por diante. Suas potências também são rotações, já que  $\mathcal{T}\mathcal{T}(x) = \mathcal{T}(\mathcal{T}x) = \mathcal{T}y$  e sua inversa é uma rotação no sentido oposto. O vetor  $\mathcal{T}x$  tem entradas dadas por  $[\mathcal{T}x]_j = x_{(j-1) \bmod N}$  para  $j = 0, 1, \dots, N - 1$

Comparando com (3.3c) conclui-se que além de serem ortogonais com suas rotações dois a dois, as wavelets também são ortogonais com as rotações dois a dois de sua função escala. Disso é direto ver que  $x = \mathbf{P}^T \mathbf{P}x$ , explicitamente:

$$x = \mathbf{P}^T \mathbf{P}x = \begin{bmatrix} g_1 & g_3 & \dots & w_1 & w_3 & \dots \\ g_0 & g_2 & \dots & w_0 & w_2 & \dots \\ g_{N-1} & g_1 & \dots & w_{N-1} & w_1 & \dots \\ g_{N-2} & g_0 & \dots & w_{N-2} & w_0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_2 & g_4 & \dots & w_2 & w_4 & \dots \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{d} \end{bmatrix} \quad (3.14a)$$

que pode ser escrito como uma soma vetorial

$$x = \mathbf{a}_0[\mathcal{T}^1g]^T + \mathbf{a}_1[\mathcal{T}^3g]^T \dots + \mathbf{d}_0[\mathcal{T}^1w]^T + \mathbf{d}_1[\mathcal{T}^3w]^T \dots \quad (3.14b)$$

---

<sup>3</sup>O mesmo resultado também vale para  $L > N$ , mas a prova é diferente. Ver Percival e Walden [46, p. 77]

Escrevendo  $[\mathcal{T}^j g]^T = \mathbf{g}_j$  chamados de funções (vetores) escala e  $[\mathcal{T}^j w]^T = \mathbf{w}_j$ , chamadas de funções (vetores) wavelet. Juntas elas formam um conjunto de  $N$  vetores linearmente independentes e portanto formam uma base para  $\mathbb{R}^N$ . Cada  $\mathbf{w}_j$  é uma função rotacionada (para baixo) de  $\mathbf{w}_0$ , e é identificado pelas coordenadas da forma  $x_{2^l+1}$ <sup>4</sup> chamadas de coordenadas diádicas. Da mesma forma os vetores  $\mathbf{g}$  são rotações de  $\mathbf{g}_0$ . Agrupando os termos em (3.14) em duas partes, uma devida às funções escala ( $\mathbf{S}$ ) e a outra devida às funções wavelets ( $\mathbf{D}$ ), obtemos o que se chama uma análise multi resolução de nível 1 (MRA, do inglês *multi resolution analysis*). Assim escreve-se (3.14) como

$$x = \mathbf{S} + \mathbf{D} \quad (3.15)$$

De grosso modo, o termo  $\mathbf{S}$  (de *smooth*) representa a parte suavizada de um sinal: picos e vales tendem a ser menores. É a parte do sinal que é identificada como tendências ou oscilações longas. Já  $\mathbf{D}$  (de *detail*) representa todas as oscilações curtas e rápidas, como mudanças estruturais ou ruídos. Como cada  $\mathbf{g}_j$  e cada  $\mathbf{w}_j$  tem norma 1, eles são ortogonais entre si, então pelo Teorema de Parseval,  $\|x\|_2 = \|\mathbf{S}\|_2 + \|\mathbf{D}\|_2$ . Isto implica que se um sinal variar pouco a cada dois intervalos de observação,  $\|\mathbf{D}\|_2$  vai ser próximo de zero. Mas isso é uma afirmação vaga, afinal o que é variar pouco? Na verdade, cada wavelet da família das Daubechies tem uma propriedade que faz dela muito útil na hora de separar o que é uma tendência do que é um detalhe. Se  $w$  é da família das Daubechies e tem suporte  $L$ , e  $x$  puder ser aproximado por um polinômio de grau até  $\frac{L}{2}$ , então os valores de  $\mathbf{d}$  serão iguais a zero, com exceção talvez de alguns poucos deles [58]. Essa última afirmação é devida ao fato de que se um sinal não for contínuo, mas cada lado de suas descontinuidades puder ser aproximado por um polinômio, então o valor de  $\mathbf{d}$  identificado com aquela coordenada, não vai ser igual a zero. Isso também só vale se o sinal não tiver termo de erro aditivo, mas se o erro tiver média zero, o sinal puder ser aproximado por um polinômio de grau até  $\frac{L}{2}$ , então  $\mathbf{d}$  terá média zero [17].

Como foi visto, em seu primeiro nível, a DWT considera mudanças ocorridas no sinal com *updates* no cálculo dos coeficientes de dois em dois. Os níveis seguintes consideram coeficientes a partir de valores espaçados por  $2^j$  unidades<sup>5</sup>. Assim, identifica-

---

<sup>4</sup>Isso por que começamos a numerar as coordenadas dos vetores por zero em vez de um

<sup>5</sup>O nível um é calculado avançando duas coordenadas, o nível dois avançando quatro, o nível três oito e assim por diante.



se cada nível da transformada com uma escala dada por uma potência de dois. Mas em vez de aplicar diretamente em  $x$ , os níveis seguintes da transformada wavelet são aplicados apenas na parte que se refere aos coeficientes *average* do nível anterior. Seguindo a notação de Walker [58], a partir de agora, para controlar o nível do qual falamos, colocamos o sobre escrito  $j$  tanto nos vetores  $\mathbf{a}$  e  $\mathbf{d}$  quanto na matriz  $\mathbf{P}$  (e.g.  $\mathbf{a}^2$  é o vetor dos coeficientes *average* do segundo nível). Como  $\mathbf{a}^1$  tem  $\frac{N}{2}$  valores, para aplicar a DWT são necessários vetores  $w$  e  $g$  estendidos para o tamanho certo. Se  $L \leq \frac{N}{2}$  então isso é feito simplesmente removendo-se zeros entre  $w_0$  e  $w_{N-1}$ . No caso de Haar para  $N = 8$  :

$$\left( \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0, 0, 0, 0, 0 \right) \longrightarrow \left( \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0 \right).$$

Os vetores análogos a (3.4) e (3.10) com metade do tamanho são  $\mathbf{W}^2$  e  $\mathbf{G}^2$ . Da mesma forma que foi obtida a primeira matriz  $\mathbf{P}^1$  é obtida a matriz

$$\mathbf{P}^2 = \begin{bmatrix} \mathbf{G}^2 \\ \mathbf{W}^2 \end{bmatrix},$$

agora de dimensão  $\frac{N}{2} \times \frac{N}{2}$ . Pelos mesmos motivos já apontados ela também é ortogonal e  $(\mathbf{P}^2)^T \mathbf{P}^2 = I_{\frac{N}{2}}$ . Assim

$$\mathbf{P}^2 \mathbf{a}^1 = \begin{bmatrix} \mathbf{a}^2 \\ \mathbf{d}^2 \end{bmatrix},$$

onde  $\mathbf{a}^2$  pode ser visto como uma média de médias e  $\mathbf{d}^2$  como uma diferença entre as médias.

Para ver por que pode-se dizer que a nível dois tem escala quatro ( $2^2$ ) observamos que as entradas de  $\mathbf{P}^2 \mathbf{a}^1$  são somas de  $L$  valores de  $\mathbf{a}^1$  consecutivos e a soma seguinte exclui os dois últimos destes valores e acrescenta os dois valores seguintes. Para o cálculo de duas entradas adjacentes de  $\mathbf{a}^1$  foram excluídas/incluídas duas entradas de  $x$ , e para entradas que estão a uma distância de dois, foram excluídas/incluídas quatro entradas de  $x$ . Já que o nível dois da transformada considera valores espaçados por duas entradas, se em vez de aplicar a transformada em  $\mathbf{a}^1$ , ela fosse aplicada em  $x$ , seriam necessários

valores de  $x$  de quatro em quatro. Isso é fácil de verificar no caso de Haar:

$$\mathbf{a}_0^1 = \frac{x_1 + x_0}{\sqrt{2}}, \quad \mathbf{a}_1^1 = \frac{x_3 + x_2}{\sqrt{2}}, \quad \mathbf{a}_2^1 = \frac{x_5 + x_4}{\sqrt{2}}, \quad \mathbf{a}_3^1 = \frac{x_7 + x_6}{\sqrt{2}}$$

$$\mathbf{a}_0^2 = \frac{\mathbf{a}_1 + \mathbf{a}_0}{\sqrt{2}}, \quad \mathbf{a}_1^2 = \frac{\mathbf{a}_3 + \mathbf{a}_2}{\sqrt{2}}.$$

desenvolvendo,

$$\mathbf{a}_0^2 = \frac{\frac{x_0+x_1}{\sqrt{2}} + \frac{x_3+x_2}{\sqrt{2}}}{\sqrt{2}} = \frac{x_0 + x_1 + x_2 + x_3}{2}$$

$$\mathbf{a}_1^2 = \frac{\frac{x_4+x_5}{\sqrt{2}} + \frac{x_6+x_7}{\sqrt{2}}}{\sqrt{2}} = \frac{x_4 + x_5 + x_6 + x_7}{2}$$

que como se vê, são somas de valores do sinal de quatro em quatro. Um argumento semelhante mostra que os níveis superiores utilizam valores do sinal espaçados por  $2^j$ .

Broughton e Bryan [9] fazem uma modificação na matriz  $\mathbf{P}^2$  para que ela possa ser aplicada diretamente em  $\mathbf{P}^1 x$ . Como o segundo nível age somente em  $\mathbf{a}^2$  e este é a primeira metade do vetor coluna dado pela DWT do primeiro nível, se completarmos  $\mathbf{P}^2$  da seguinte forma:

$$\mathbf{P}^{2*} = \begin{bmatrix} \mathbf{P}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\frac{N}{2}} \end{bmatrix} \quad (3.16)$$

onde  $\mathbf{0}$  é um bloco de zeros de forma que teremos uma matriz ortogonal<sup>6</sup> de dimensão  $N \times N$ . De agora em diante escreveremos  $\mathbf{P}^2$  como a matriz em (3.16). Fazendo o produto  $\mathbf{P}^2 \mathbf{P}^1 x$  obtemos

$$\begin{bmatrix} \mathbf{a}^2 \\ \mathbf{d}^2 \\ \mathbf{d}^1 \end{bmatrix}$$

e pelo fato de as matrizes serem ortogonais, o produto delas também é uma matriz ortogonal e sua transposta é dada por  $(\mathbf{P}^1)^T (\mathbf{P}^2)^T = \mathbf{P}^*$  e assim

$$x = \mathbf{P}^* \mathbf{P}^2 \mathbf{P}^1 x = \mathbf{P}^* \begin{bmatrix} \mathbf{a}^2 \\ \mathbf{d}^2 \\ \mathbf{d}^1 \end{bmatrix},$$

---

<sup>6</sup>Para ver isso, basta utilizar a multiplicação de matrizes em blocos

que pode ser escrito como em (3.14):

$$x = \mathbf{a}_0^2 \mathbf{g}_0^2 + \dots + \mathbf{a}_{\frac{N}{4}}^2 \mathbf{g}_{\frac{N}{4}}^2 + \mathbf{d}_0^2 \mathbf{w}_0^2 + \dots + \mathbf{d}_{\frac{N}{4}}^2 \mathbf{w}_{\frac{N}{4}}^2 + \mathbf{d}_0^1 \mathbf{w}_0^1 + \dots + \mathbf{d}_{\frac{N}{2}}^1 \mathbf{w}_{\frac{N}{2}}^1 \quad (3.17)$$

Os vetores  $\mathbf{w}_0^1 + \dots + \mathbf{w}_{\frac{N}{2}}^1$  já são conhecidos, pois por construção a metade da direita  $\mathbf{P}^*$  é  $(\mathbf{W}^1)^T$  (eles são os mesmos que no primeiro nível). Já os vetores  $\mathbf{g}_j^2$  e  $\mathbf{w}_j^2$  são obtidos da seguinte forma: escrevendo explicitamente  $\mathbf{P}^*$  como um produto de matrizes

$$\mathbf{P}^* = \underbrace{\begin{bmatrix} g_1 & g_3 & \dots & w_1 & w_3 & \dots \\ g_0 & g_2 & \dots & w_0 & w_2 & \dots \\ g_{N-1} & g_1 & \dots & w_{N-1} & w_1 & \dots \\ g_{N-2} & g_0 & \dots & w_{N-2} & w_0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_2 & g_4 & \dots & w_2 & w_4 & \dots \end{bmatrix}}_{(\mathbf{P}^1)^T} \underbrace{\begin{bmatrix} g_1 & g_3 & \dots & w_1 & w_3 & 0 & \dots \\ g_0 & g_2 & \dots & w_0 & w_2 & 0 & \dots \\ g_{N-1} & g_1 & \dots & w_{N-1} & w_1 & 0 & \dots \\ g_{N-2} & g_0 & \dots & w_{N-2} & w_0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots \end{bmatrix}}_{(\mathbf{P}^2)^T} \quad (3.18)$$

calculando a primeira e a  $\frac{N}{4}$ -ésima colunas de  $\mathbf{P}^*$  temos, desde que  $L \leq \frac{N}{2}$ ,

$$\begin{aligned} [\mathbf{P}^*]_{1,1} &= g_1^2 + g_0 g_3 + g_{N-1} g_5 + \dots \\ [\mathbf{P}^*]_{2,1} &= g_1 g_0 + g_0 g_2 + g_{N-1} g_4 + \dots \\ [\mathbf{P}^*]_{3,1} &= g_1 g_{N-1} + g_0 g_1 + g_{N-1} g_3 + \dots \\ [\mathbf{P}^*]_{3,1} &= g_1 g_{N-2} + g_0^2 + g_{N-1} g_2 + \dots \\ &\vdots \end{aligned}$$

e

$$\begin{aligned} [\mathbf{P}^*]_{1,\frac{N}{4}} &= w_1 g_1 + w_0 g_3 + w_{N-1} g_5 + \dots \\ [\mathbf{P}^*]_{2,\frac{N}{4}} &= w_1 g_0 + w_0 g_2 + w_{N-1} g_4 + \dots \\ [\mathbf{P}^*]_{3,\frac{N}{4}} &= w_1 g_{N-1} + w_0 g_1 + w_{N-1} g_3 + \dots \\ [\mathbf{P}^*]_{3,\frac{N}{4}} &= w_1 g_{N-2} + w_0 g_0 + w_{N-1} g_2 \\ &\vdots \end{aligned}$$

Esse padrão se repete para as outras colunas de  $\mathbf{P}^*$  até a de número  $\frac{N}{2}$ . Além disso, olhando atentamente para ele, percebe-se que em cada multiplicação os termos

da matriz à direita aparecem primeiro. Isso foi proposital, pois assim percebemos uma recorrência. Cada uma das colunas da matriz  $\mathbf{P}^*$  é formada como uma combinação linear dos vetores coluna  $\mathbf{g}_0^1$  e suas rotações. Por exemplo:

$$\mathbf{g}_0^2 = g_1 \mathbf{g}_0^1 + g_0 \mathbf{g}_1^1 + g_{N-1} \mathbf{g}_2^1 \dots$$

e

$$\mathbf{w}_0^2 = w_1 \mathbf{g}_0^1 + w_0 \mathbf{g}_1^1 + w_{N-1} \mathbf{g}_2^1 \dots$$

Tal qual o primeiro nível, as outras colunas podem ser encontradas com rotações destas duas primeiras. Nota-se que as entradas diferentes de zero desses vetores não é mais  $L$  e sim  $3(L-1) + 1$  [46]. Os níveis seguintes da DWT são construídos da mesma maneira, mas agora a matriz equivalente a (3.16) para o nível  $j$  tem um bloco de dimensão  $\frac{N}{2^j} \times \frac{N}{2^j}$  a direita e acima, a diagonal completa com uns e o resto completo por blocos de zeros. Contudo que  $(2^j - 1)(L - 1) + 1 \leq \frac{N}{2^j}$  <sup>7</sup> as  $\frac{N}{2^j}$  primeiras colunas dessa matriz são dadas por

$$\mathbf{g}_0^j = g_0 \mathbf{g}_0^{j-1} + g_1 \mathbf{g}_2^{j-1} + g_{N-1} \mathbf{g}_2^{j-1} \dots$$

e

$$\mathbf{w}_0^{2^j} = w_0 \mathbf{g}_0^{j-1} + w_1 \mathbf{g}_2^{j-1} + w_{N-1} \mathbf{g}_2^{j-1} \dots$$

Um ponto importante que deve ser enfatizado é que sempre foi feita a hipótese de que o suporte das wavelets fosse menor que o tamanho do vetor no qual a transformada é aplicada. Isso não é necessário (para nenhum nível) pois sempre é feita a hipótese de que o vetor é periódico, portanto, se  $L$  foi maior do que  $N$ , por meio de extensão periódica podemos aumentar o vetor para satisfazer o requerimento. Computacionalmente isso é o mesmo que somar os coeficientes de  $g$  ou  $w$  que multiplicam valores de  $x$  que foram estendidos, mantendo a matriz da transformada com tamanho  $N$ . Porém isso tem um custo, já que não há nenhum motivo para que um sinal qualquer seja realmente periódico, e, de fato, pelo menos com dados estatísticos, eles tendem a não ser. Por isso deve se tomar cuidado com o tamanho do suporte das wavelets que se usam. Formalmente, a utilização de apenas alguns níveis da transformada visando não fazer uso excessivo da

---

<sup>7</sup>O tamanho dos vetores que compõem as  $\frac{N}{2^j}$  primeiras colunas da matriz que representa a DWT tem suporte dado por  $(2^j - 1)(L - 1) + 1$  [46]

hipótese de periodicidade chama-se **Transformada Wavelet Parcial** [46], mas aqui não será feita essa distinção. Na verdade nem será formalizado o que é a transformada wavelet completa em vez da parcial. Em vez disso, dado um sinal e uma wavelet, serão feitas quantas transformações forem adequadas dados o tamanho do sinal e o suporte da wavelet,  $w$ , que deve ter as propriedades úteis ao problema que se deseja resolver. Ainda assim é necessário que  $N$  seja divisível por dois pelo menos  $j$  vezes para realizar  $j$  níveis da transformada, isso por que cada nível sempre requer metade do número de valores do nível anterior. Assim é feito o uso mínimo da hipótese de periodicidade de um sinal.

Assim como as equações (3.14) e (3.15) para o primeiro nível, os outros níveis da transformada também podem ser escritos em formato de MRA.

$$\begin{aligned}
 x = & \underbrace{\sum_{i=0}^{\frac{N}{2^J}-1} \mathbf{a}_i^J \mathbf{g}_i^J}_{\mathbf{S}^J} + \underbrace{\sum_{i=0}^{\frac{N}{2^J}-1} \mathbf{d}_i^J \mathbf{w}_i^J}_{\mathbf{D}^J} + \underbrace{\sum_{i=0}^{\frac{N}{2^{J-1}}-1} \mathbf{d}_i^{J-1} \mathbf{w}_i^{J-1}}_{\mathbf{D}^{J-1}} + \underbrace{\sum_{i=0}^{\frac{N}{2^{J-2}}-1} \mathbf{d}_i^{J-2} \mathbf{w}_i^{J-2}}_{\mathbf{D}^{J-2}} \\
 & + \dots + \underbrace{\sum_{i=0}^{\frac{N}{2}-1} \mathbf{d}_i^1 \mathbf{w}_i^1}_{\mathbf{D}^1}
 \end{aligned} \tag{3.19}$$

Começando com o nível dois  $x = \mathbf{S}^2 + \mathbf{D}^2 + \mathbf{D}^1$  e comparando com (3.15) chega-se a conclusão que  $\mathbf{S}^1 = \mathbf{D}^2 + \mathbf{S}^2$ , o que deveria se esperar, pois calculamos o segundo nível baseado nos coeficientes *average* do nível um. Em geral  $\mathbf{S}^{J-1} = \mathbf{D}^J + \mathbf{S}^J$  [46]. Isto é, a transformada wavelet produz em cada nível uma aproximação de  $x$  por meio de seus valores médios e uma diferença, e um nível imediatamente superior age na aproximação do nível abaixo e produz uma nova aproximação (menos exata) e um novo detalhe. Realizando estas trocas podemos decompor um sinal do jeito que quisermos, e como sabemos o tamanho de cada escala, identificamos cada nível com um instante conhecido.

Para dar um exemplo, realizamos a transformada wavelet em um sinal simulado dado por  $x = 0.1t + \varepsilon_t$ , onde  $\varepsilon_t$  é um termo aleatório obtido como comando `rand` do Scilab. A taxa de amostragem é  $\frac{1}{128}$  então, sem normalizar o intervalo,  $t$  varia de zero a 127. Como este sinal tem um termo linear, para os coeficientes detalhe terem média zero, utiliza-se uma wavelet da família das Daubechies que tenha suporte  $L = 4$ . Essa wavelet

é chamada DB2 e tem os coeficientes dados por :

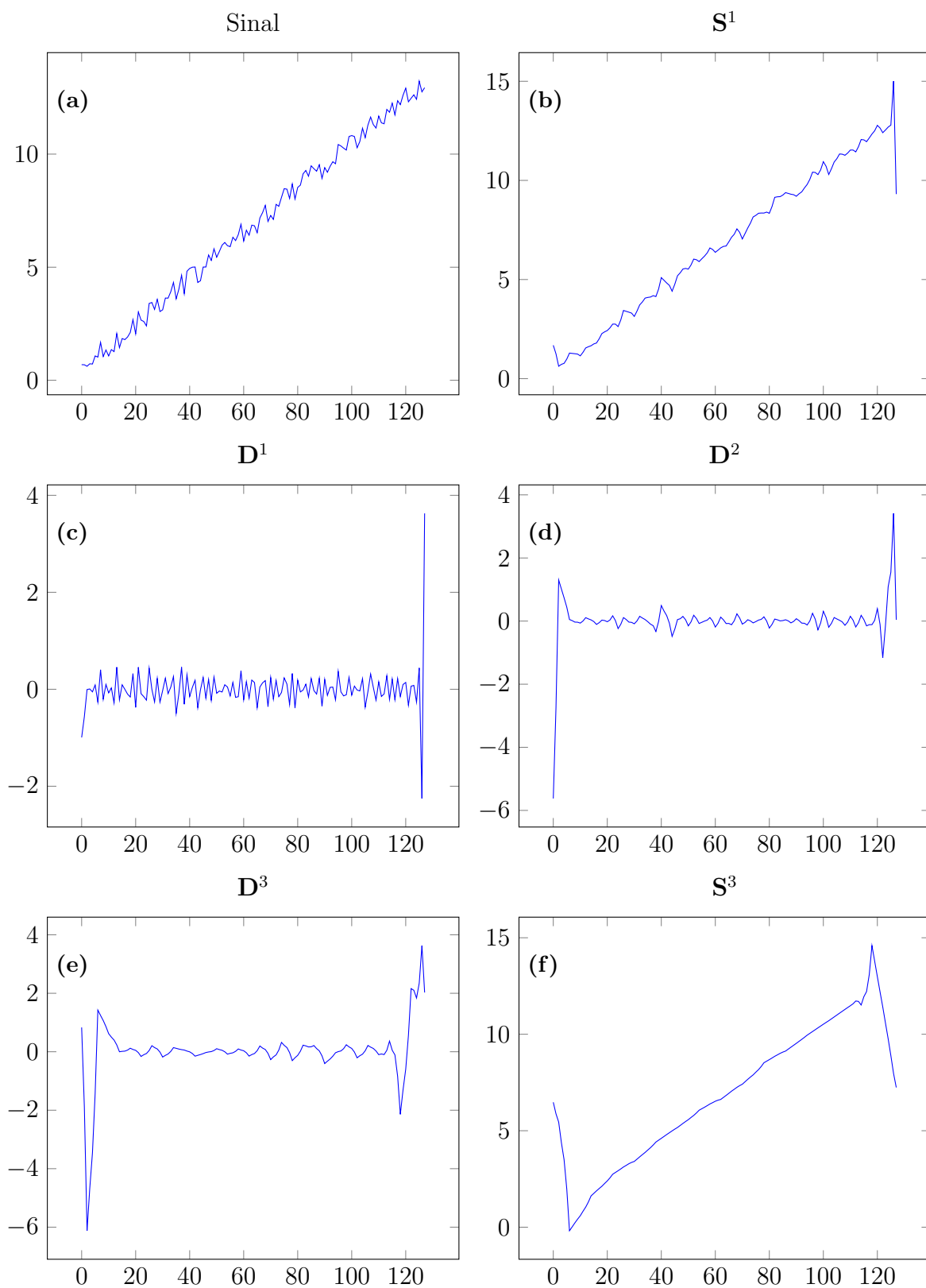
$$g = (-0.1294, 0.2241, 0.8365, 0.4829) \quad (3.20a)$$

$$w = (-0.4829, 0.8365, -0.2241, -0.1294). \quad (3.20b)$$

Os coeficientes acima são aproximações com quatro casas decimais dos coeficientes da DB2 disponíveis no website Wavelet Browser [59], que dispõe de uma grande biblioteca de wavelets, com os coeficientes já calculados e em formato decimal. A figura 3.1 apresenta os gráficos da Análise Multirresolução aplicada nesse sinal simulado.

**Definição 9.** A decomposição de um vetor  $x$  em colunas formadas com a transposição das linhas da DWT é chamada de análise multirresolução. Este nome é devido ao fato de que podemos trocar sempre níveis *average* mais baixos, que representam a tendência do sinal em intervalos (escalas) curtos de tempo por um componente *detail*, de oscilações, nestes intervalos mais um componente *average* de um nível superior, que representa intervalos de tempo maiores:

**Figura 3.1:** Exemplo DB2



Fonte: elaborado pelo autor. O eixo das abscissas representa o tempo.

Como se pode ver, os componentes *smooth* (Figura 3.1b e 3.1f) captam, aproximadamente, a tendência geral do sinal de crescer com o tempo enquanto que os componentes *detail* (Figuras 3.1c, 3.1d e 3.1e) captam os desvios desse sinal. Além disso, nos termos *smooth*, observamos duas coisas: no interior do gráfico eles apresentam um formato serrilhado – isso é devido ao termo de ruído que polui as médias, embora ele seja suavizado, ele não é removido; nas pontas eles apresentam grandes saltos – isso é devido a hipótese de periodicidade do sinal. Se ele fosse estendido para além de 128 amostras para ambos os lados (i.e. incluindo tempos negativos) essas discontinuidades desapareceriam nesses locais, mas reapareceriam nas novas pontas. Como resposta, os termos *detail* também apresentam essas discontinuidades. Feita essa ressalva, notamos que a aproximação obtida por  $\mathbf{S}^3$  é quase uma linha reta e capta quase toda a informação sobre a tendência do sinal. Comparando com o exemplo do filtro de Haar, vemos que ambas metodologias podem ser usadas para limpar um sinal de termos de grande oscilação em torno de uma tendência que oscila relativamente mais devagar.

Como para cada nível, a matriz  $\mathbf{P}^{J*}$  similar a (3.16), com os blocos dimensionados de acordo com o nível, é ortogonal, e o como produto de matrizes ortogonais, também é ortogonal pelo Teorema de Parseval, os vetores  $x$  e  $y = ([\mathbf{a}^J]^T, [\mathbf{d}^J]^T, [\mathbf{d}^{J-1}]^T, \dots, [\mathbf{d}^1]^T)^T$  são tais que:

$$\|x\|_2^2 = \|y\|_2^2 = \|\mathbf{a}^J\|_2^2 + \|\mathbf{d}^J\|_2^2 + \|\mathbf{d}^{J-1}\|_2^2 + \dots + \|\mathbf{d}^1\|_2^2 \quad (3.21)$$

Onde cada norma do lado direito de (3.21) é tomada de acordo com o tamanho dos sub vetores que compõem  $y$ , já que em geral eles não tem o mesmo tamanho <sup>8</sup>. Para cada escala  $2^j = \tau_j$  a função

$$\mathcal{P}_{\tau_j}(x) = \begin{cases} \|\mathbf{d}^j\|_2, & \text{se } j < J \\ \|\mathbf{a}^{j-1}\|_2 & \text{se } j = J \text{ e } J > 0 \\ \|x\|_2, & \text{se } j = 0 \end{cases} \quad (3.22)$$

é chamada de escalômetro [46]. Analogamente ao espectrômetro, ela decompõe a energia de um sinal em componentes, no caso em escalas, i.e. quanto cada escala representa para

---

<sup>8</sup>Isso pode ser feito já que a norma ao quadrado é simplesmente um somatório de valores independente da dimensão do espaço.



a energia total de um sinal. No caso da transformada de Fourier, era possível escolher algumas frequências e montar um sinal apenas com elas, aceitando um erro de estimação, ou seja, que o vetor expurgado de certas frequências seja apenas próximo do original, segundo a equação (2.10). O mesmo pode ser feito com o escalômetro e com a DWT.

## 3.2 Wavelets como Filtros Lineares

Na seção anterior foi apresentada a DWT como se fosse uma transformada ortogonal qualquer. Isso é conveniente por que é simples de se fazer e fácil de se implementar, porém tem uma grande desvantagem: se um vetor é rotacionado, o resultado da DWT é modificado. Isto é um problema em aplicações estatísticas. Uma metodologia que oferece respostas diferentes dependendo dos pontos iniciais, mesmo que eles sejam próximos, é uma metodologia potencialmente inconsistente. Felizmente existe uma pequena mudança que se pode fazer na DWT para que ela seja insensível a rotações. Essa é a proposta de Percival e Walden [46] com a *Maximum Overlap Discret Wavelet Transform*, ou MODWT, que parte da DWT e constrói o que talvez seja a mais importante ferramenta para aplicação de wavelets às séries temporais.

Tratando-se da transformada de Fourier, Walker [58] mostra, com o uso da transformada  $Z$ , que o espectro do vetor  $\mathcal{T}x$  é, entrada a entrada  $e^{\frac{-2\pi ik}{N}} \mathbf{X}_k$ , que pela discussão envolvendo a equação (2.22), é uma mudança de fase dos coeficientes do vetor  $x$ . Portanto após uma rotação mantém-se o mesmo espectro e apenas atrasa-se ou adianta-se cada frequência e, contanto que o sinal seja periódico, não importa em qual o período é feita uma amostragem seu espectro não vai mudar.

Embora tenha sido feita a hipótese de que o sinal é periódico anteriormente e até mostrado que a descontinuidade induzida ao estender um vetor periodicamente cria discrepâncias na MRA obtida por uma wavelet, pelo menos para o cálculo dos coeficientes, uma grande porção deles não precisa da hipótese de periodicidade. Em geral ela só é necessária nas  $(2^j - 1)(L - 1) + 1$  primeiras entradas, que são a quantidade de valores do sinal utilizados para o cálculo dos coeficientes, em suas respectivas escalas. Isso significa que a transformada wavelet consegue decompor, aproximadamente, a energia de um sinal em escalas e manter essa decomposição confiável quando novas informações entram, exceto

nos primeiros coeficientes de cada nível. Isso obviamente é muito útil em aplicações a séries temporais que são atualizadas periodicamente. Com wavelets, até certo ponto, não é necessária que a hipótese de periodicidade seja feita de forma enfática.

Esse ganho comparativo porém é diminuído quando um sinal, em vez de ser incrementado, tem observações feitas, em mesmo número, mas em datas diferentes. Por exemplo, começar um experimento de uma observação por dia na segunda-feira ou na terça-feira. Idealmente isso não deveria ter nenhum efeito sobre a MRA de  $x$ . Componentes avaliados de dois em dois dias (primeira escala), quatro em quatro (segunda escala) etc., deveriam contribuir igualmente com a energia de  $x$  seja qual for a data inicial do experimento. Porém isso não ocorre. Por exemplo o sinal  $x = (1, 1, -1, -1)$  se rotacionado uma vez fica  $x' = (-1, 1, 1, -1)$ . Aplicando a função escala e a wavelet de Haar nesses sinais obtemos  $\mathbf{a}^1 = (\sqrt{2}, -\sqrt{2})$  e  $\mathbf{d}^1 = (0, 0)$  para  $x$  e  $\mathbf{a}'^1 = (0, 0)$   $\mathbf{d}'^1 = (\sqrt{2}, -\sqrt{2})$ . Claramente a rotação fez diferença. Por outro lado se fossem feitas duas rotações os coeficientes  $\mathbf{a}$  e  $\mathbf{d}$  manteriam seus valores, exceto por uma rotação. Em resumo, a DWT é indiferente apenas a pares de rotações. Visando eliminar esse problema Percival e Walden [46] mostram a *Maximum Overlap Discret Wavelet Transform* que é indiferente a qualquer número de rotações, a menos de uma rotação dela mesmo. Eles fazem essa construção estudando as wavelets como filtros lineares.

O filtro wavelet  $w$  contínua satisfazendo as condições em (3.3) e ele pode ser utilizado na forma (2.15), reproduzida a seguir

$$W = \begin{bmatrix} w_0 & w_{N-1} & w_{N-2} & \dots & w_2 & w_1 \\ w_1 & w_0 & w_{N-1} & \dots & w_3 & w_2 \\ \vdots & & \ddots & & & \vdots \\ w_{N-1} & w_{N-2} & w_{N-3} & \dots & w_1 & w_0 \end{bmatrix}$$

A condição (3.3a) diz que  $w$  é um filtro passa-alto pois garante que o produto interno de  $w$  e suas rotações aplicado a um vetor constante é sempre igual a zero. Se  $\mathbf{W}_k$  são os valores da DFT de  $w$  então pelo Teorema da Convolução,  $\mathbf{d} = Wx$  tem a DFT dada por  $\mathbf{W}_k \mathbf{X}_k$ . Junto de  $w$  temos a função escala  $g$  com DTF dada entrada a entrada por  $\mathbf{G}_k$ . Juntos eles formam um par de filtros de quadratura [46], logo se  $\mathcal{W}$  e  $\mathcal{G}$  são as funções ganho de  $w$  e  $g$  respectivamente, então  $\mathcal{G}(k) + \mathcal{W}(k) = C$ , para todo  $k$  e uma constante  $C$ . Por construção, Percival e Walden [46, p. 70] estabelecem que essa constante é igual

a 2 e já que  $\mathcal{W}(0) = 0$  temos necessariamente que  $\mathcal{G}(0) = 2$  e isso só acontece se

$$(g_0 + g_1 + \dots + g_{L-1})^2 = 2$$

portanto

$$g_0 + g_1 + \dots + g_{L-1} = \sqrt{2}.$$

que é uma das propriedades da função escala já comentada.

Se os vetores  $g$  e  $w$  forem multiplicados por  $\frac{1}{\sqrt{2}}$ , obtendo-se  $\tilde{g}$  e  $\tilde{w}$ , e a constante  $C$  se torna 1. Então o vetor obtido somando-se  $(x * \tilde{w})$  e  $(x * \tilde{g})$  têm o mesmo conteúdo de frequência que  $x$ , enquanto que a soma de  $(x * w)$  e  $(x * g)$  tem conteúdo de frequência igual a duas vezes o de  $x$ .

Para estabelecer a relação entre a DWT e os filtros wavelet Broughton e Bryan [9] definem o operador de subamostragem  $D$ .

**Definição 10.** O operador de sub amostragem  $D : (y_0, y_1, \dots, y_{N-1}) \mapsto (y_1, y_3, \dots, y_{N-1})$  elimina as entradas pares de um vetor  $x$  e constrói um vetor novo contendo apenas as coordenadas ímpares.

Se  $x_h$  for o vetor resultante em aplicar a matriz de convolução  $W$  a  $x$  então  $D(x_h)$  é dado entrada a, entrada por

$$[D(x_h)]_k = \langle \mathcal{T}^{2k+1}w, x \rangle. \quad (3.23)$$

Mas esses valores são exatamente os  $\frac{N}{2}$  valores calculados pela DWT referentes ao vetor  $\mathbf{d}^1$  da seção anterior. Se  $G$  é a matriz de convolução gerada pelo vetor  $g$  e  $x_l = Gx$ , e  $D(x_l)$  é dado entrada a entrada por

$$[D(x_l)]_k = \langle \mathcal{T}^{2k+1}g, x \rangle, \quad (3.24)$$

que são exatamente os mesmos valores referentes ao vetor  $\mathbf{a}^1$  da seção anterior. Portanto o primeiro nível da DWT é resultante de uma operação de filtragem, seguida de uma operação de subamostragem. Se executadas ao mesmo tempo, temos a Transformada Wavelet Ortogonal [9], que foi apresentada na seção anterior. Obviamente poderíamos realizar as duas operações em separado, mas isso resultaria em cálculos redundantes.

Agora, utilizar as ferramentas de filtros para provar que é possível inverter a operação de filtragem seguida de subamostragem é uma tarefa um pouco mais delicada. Primeiro por que nem mesmo a operação de filtragem é invertível, já que  $W$  e  $G$  têm autovalores nulos. Segundo, subamostrar um vetor ocasiona ainda mais perda de informação. Do ponto de vista da recuperação de um sinal isso dificulta ainda mais as coisas. Broughton e Bryan [9] mostram que com um banco de filtros, i.e. usar filtros  $p_a$  e  $p_b$  apropriados em conjunto, pode-se recuperar o sinal completo após filtragem e subamostragem através de uma nova filtragem com o que eles chama de filtro de síntese.

Como neste trabalho serão utilizados apenas filtros prontos, o que é importante dos filtros síntese é a operação de sobre-amostragem.

**Definição 11.** O operador de sobre-amostragem  $U(y) : (y_0, y_1, \dots, y_{N-1}) \mapsto (0, y_0, 0, y_1, \dots, 0, y_{N-1})$  insere zeros antes de cada coordenada de um vetor, dobrando seu tamanho.

Para não ter de falar sobre filtros de síntese, Pericival e Walden [46] apenas verificam o que acontece quando se faz o produto  $W^T[U(D(x_h))]$ , explicitamente:

$$W^T[U(D(x_h))] = \begin{bmatrix} w_0 & w_1 & w_2 & w_3 & \dots & w_{N-1} \\ w_{N-1} & w_0 & w_1 & w_2 & \dots & w_{N-2} \\ w_{N-2} & w_{N-1} & w_0 & w_1 & \dots & w_{N-3} \\ w_{N-3} & w_{N-2} & w_{N-1} & w_0 & \dots & w_{N-4} \\ \vdots & & \ddots & & & \vdots \\ w_1 & w_2 & w_3 & w_4 & \dots & w_0 \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{d}_0 \\ 0 \\ \mathbf{d}_1 \\ \vdots \\ \mathbf{d}_{\frac{N}{2}} \end{bmatrix}, \quad (3.25)$$

onde os elementos de  $x_h$  foram escritos como os elementos do vetor  $\mathbf{d}^1$  da seção anterior. Se chamarmos as colunas da matriz do lado direito de (3.25) de  $\mathbf{c}_1, \mathbf{c}_2, \dots$  então temos

$$W[U(D(x_h))] = 0\mathbf{c}_1 + \mathbf{d}_0\mathbf{c}_2 + 0\mathbf{c}_3 + \mathbf{d}_1\mathbf{c}_4 \dots$$

Comparando com a matriz em (3.14), vemos que as colunas com índice par são as mesmas da segunda metade da matriz  $\mathbf{P}^T$ . A soma portanto equivale a somar a segunda metade das colunas de  $\mathbf{P}^T$  com pesos dados pelo vetor  $\mathbf{d}$ .

$$W^T[U(D(x_h))] = \mathbf{d}_0 W_0^T + \mathbf{d}_1 W_1^T + \mathbf{d}_2 W_2^T \dots$$

Seguindo o mesmo raciocínio, a convolução seguida de sub e sobre-amostragem com  $g = G^T[U(D(x_l))]$  resulta em:

$$G^T[U(D(x_l))] = \mathbf{a}_0 G_0^T + \mathbf{a}_1 G_1^T + \mathbf{a}_2 G_2^T \dots$$

que é a primeira metade das colunas da matriz  $\mathbf{P}^T$  somadas com pesos dados pelo vetor  $\mathbf{a}^1$ . Além disso o formato da matriz em (3.25) já foi visto, pois é a convolução especial que recebe o nome de correlação cruzada. Assim mostra-se que a DWT pode ser vista como um procedimento de filtragem com um banco de filtros, seguido de uma subamostragem, uma sobre-amostragem e uma correlação cruzada.

### 3.2.1 As Daubechies no domínio da frequência e os níveis superiores da DWT

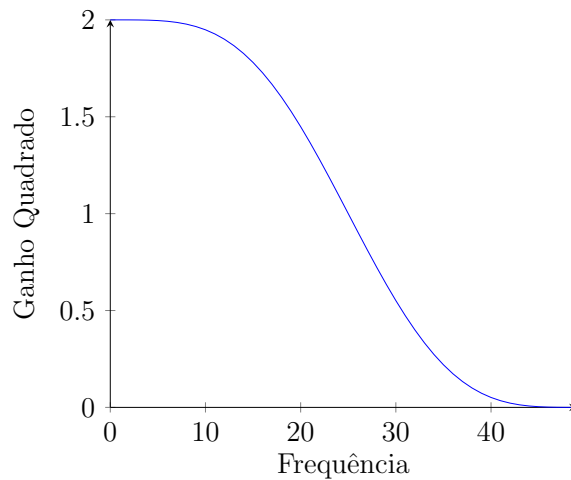
Quando foi apresentada a DWT como uma transformada linear, foi dito que cada nível representava uma escala diferente e cada coluna da matriz de síntese representava uma localização temporal específica. Em se tratando de filtros, a representação natural de se falar é no domínio da frequência. O Teorema da Convolução garante que o vetor resultante do produto  $Wx$  terá sua DFT dada por  $\mathbf{W}_k \mathbf{X}_k$ . Observando as figuras 2.4 e 2.5, que são os gráficos da função ganho quadrado da wavelet e função escala de Haar, respectivamente, nota-se que em ambos os casos, os filtros mantêm a frequência que representa metade da frequência de Nyquist, ou  $\frac{1}{4}$  de  $N$ , intacta. Quando é utilizada a wavelet de Haar, as frequências acima de  $\frac{1}{4}$  são amplificadas e as abaixo, são atenuadas. Se for utilizada a função escala, a relação é invertida. Segundo Percival e Walden [46] as Daubechies quando utilizadas como um filtro têm essa propriedade: elas dividem o espectro de um vetor em duas bandas de mesmo tamanho, então é comum identificar o vetor  $Wx$  com frequências em módulo no intervalo  $[1/4, 1/2]$  e o vetor  $Gx$  com frequências no intervalo  $[0, 1/4]$ . Evidentemente isso é apenas uma aproximação, como se pode ver nos gráficos da função ganho, que só se anulam em um ponto.

Os outros membros da família das Daubechies têm gráficos semelhantes. A função ganho quadrado  $\mathcal{G}$  de cada função escala de uma Daubechies com suporte  $L$  é dada por [46]:

$$\mathcal{G}(f) = 2 \cos^L(\pi\omega) \sum_{l=0}^{\frac{L}{2}-1} \binom{\frac{L}{2}-1+l}{l} \text{sen}^{2l}(\pi\omega). \quad (3.26)$$

Nesse caso,  $\omega$  denota frequências contínuas como no início do capítulo passado. Quando forem utilizadas wavelets discretizadas devemos substituir as frequências contínuas por  $\frac{k}{N}$  [46]. Para encontrar  $\mathcal{W}$ , calcula-se  $2 - \mathcal{G}$ .

**Figura 3.2:** Função ganho quadrado da função escala DB2



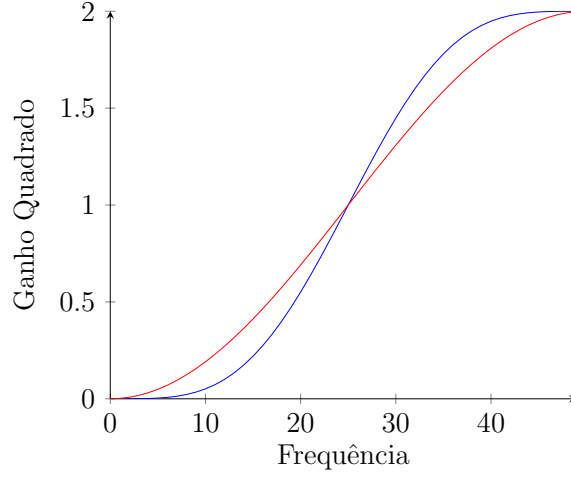
Fonte: elaborado pelo autor.

A figura 3.2 mostra  $\mathcal{G}$  para a db2. O formato é semelhante a figura 2.5. A figura 3.3 mostra  $\mathcal{W}$  da db2 (azul) e de Haar (vermelho). Nota-se que a db2 tem uma função ganho mais inclinada, isso significa que ela age melhor como um filtro de banda. Em geral, quanto maior o suporte da wavelet, melhor ela será como um filtro de banda [46]. Em contrapartida, ela vai precisar de mais dados, o que inviabiliza wavelets muito grandes para alguns problemas.

O segundo e os níveis superiores da transformada também são filtros lineares, mas eles precisam de uma argumentação um pouco diferente. Aqui, começamos pela conclusão: cada nível da DWT equivale a passar um sinal por um banco de filtros que divide a banda das frequências mais baixas do nível anterior em duas metades e uma operação de sub amostragem [46]. Por exemplo, segundo nível da DWT divide  $[0, 1/4]$ . em  $[0, 1/8]$ .  $[1/8, 1/4]$ . Em geral para o nível  $J$  a DWT é resultado da filtragem do espectro do sinal nas bandas  $[0, 1/2^{J+1}]$ .  $[1/2^{J+1}, 1/2^J]$ . Para provar esse fato Percival e Walden [46, p.91] notam que se for inserido um certo número  $2^m - 1$  de zeros entre as entradas de um filtro:

$$(w_0, w_1, w_2, \dots) \longrightarrow w^m = (w_0, \underbrace{0, 0, \dots}_{2^m - 1 \text{ zeros}}, w_1, \underbrace{0, 0, \dots}_{2^m - 1 \text{ zeros}}, w_2, \dots)$$

**Figura 3.3:** Função ganho quadrado da wavelet DB2



Fonte: elaborado pelo autor.

a relação entre a função ganho quadrado de  $w$  e a função ganho quadrado de  $w^m$  é

$$\mathcal{W}[(m+1)k] = \mathcal{W}^m(k).$$

Como exemplo, o segundo nível da DWT consiste em aplicar as matrizes de convolução  $G^2$  e  $W^2$ , construídas a partir dos vetores  $g^2$  e  $w^2$ , no vetor  $G^1x$ , obtendo os vetores  $x_{ll} = G^2G^1x$  e  $x_{lh} = W^2G^1x$ . A DFT de cada um, omitindo-se a função fase, é dada por:

$$(\mathbf{X}_{ll})_k = |\mathcal{G}^2(k)\mathcal{G}^1(k)|\mathbf{X}_k = |\mathcal{G}^1(2k)\mathcal{G}^1(k)|\mathbf{X}_k \quad (3.27a)$$

$$(\mathbf{X}_{lh})_k = |\mathcal{W}^2(k)\mathcal{G}^1(k)|\mathbf{X}_k = |\mathcal{W}^1(2k)\mathcal{G}^1(k)|\mathbf{X}_k. \quad (3.27b)$$

Para ver como isso se relaciona a DWT escrevemos  $x_{ll}$  como um produto de uma matriz e um vetor:

$$x_{ll} = \begin{bmatrix} g_0 & 0 & g_{N-1} & 0 & g_{N-2} & \dots & 0 \\ 0 & g_0 & 0 & g_{N-1} & 0 & \dots & g_1 \\ g_1 & 0 & g_0 & 0 & g_{N-1} & \dots & 0 \\ 0 & g_1 & 0 & g_0 & 0 & \dots & g_2 \\ \vdots & & & & \ddots & & \vdots \\ 0 & g_{N-1} & 0 & g_{N-2} & 0 & \dots & g_0 \end{bmatrix} \begin{bmatrix} (x_l)_0 \\ \mathbf{a}_0^1 \\ (x_l)_2 \\ \mathbf{a}_1^1 \\ \vdots \\ (x_l)_{N-1} \end{bmatrix} \quad (3.28)$$

Nota-se que as entradas de  $x_{ll}$  são calculadas, ou com apenas componentes que estão em  $\mathbf{a}^1$  ou apenas componentes que não estão. Segundo, tomar  $D(x_{ll})$  equivale a aplicar a matriz  $G^1$  redimensionada para ter dimensão  $\frac{N}{2} \times \frac{N}{2}$  em  $\mathbf{a}^1$ . Isso, por sua vez, equivale a executar um filtro de banda em um vetor em  $\mathbb{R}^{\frac{N}{2}}$  e como já foi mostrado, se for feita a subamostragem do vetor resultante, encontramos a decomposição *average* da DWT aplicada em  $\mathbf{a}^1$ , em outras palavras, encontra-se  $\mathbf{a}^2$ . Pelo mesmo raciocínio, trocando  $G^2$  por  $W^2$  encontra-se  $\mathbf{d}^2$  e mostra-se que o segundo nível da DWT equivale a uma subamostragem de dois filtros aplicados aos coeficientes  $\mathbf{a}^1$  [46]. Equivalentemente podemos tomar cada quarto valor da equação (3.28), isto é, subamostrar os vetores  $x_{ll}$  e  $x_{lh}$  por um fator de quatro.

Os níveis seguintes seguem o mesmo padrão: criam-se os vetores  $g^m$  e  $w^m$  com  $2^m - 1$  zeros entre cada entrada diferente de zero, e filtra-se o vetor composto pelos coeficientes *average* do nível anterior com  $g^m$  e  $w^m$  dimensionados corretamente. Isso equivale a filtrar o vetor  $y = G^{m-1}G^{m-2} \dots G^1 x$  e subamostrar o resultado por um fator de  $2^m$  [46]. Aplicações sucessivas das equações (3.27a) e (3.27b) garantem que os vetores  $\mathbf{a}^m$  e  $\mathbf{d}^m$  representam vetores, subamostrados, com conteúdo de frequências nas bandas  $[0, 1/2^{m+1}]$  e  $[1/2^{m+1}, 1/2^m]$ , respectivamente.

A inversão desse procedimento se dá de forma análoga a equação (3.25), com a matriz e o vetor à direita recebendo  $2^m - 1$  zeros entre suas entradas. Utilizando a seguinte notação:  $x_{ll} = x_{2l}, x_{lll} = x_{3l}, x_{llh} = x_{2lh}, etc..$  e  $D(D(x)) = DD = D^2, D(D(D(x))) = D^3, etc..$  Escreve-se  $x$  como:

$$\begin{aligned} x = & (G^m G^{m-1} \dots G^1)^T U^m D^m(x_{ml}) + (W^m G^{m-1} \dots G^1)^T U^m D^m(x_{(m-1)lh}) \\ & + (W^{m-1} G^{m-2} \dots G)^T U^{m-1} D^{m-1}(x_{(m-2)lh}) + \dots (W^1)^T U D(x_h) \end{aligned} \quad (3.29)$$

Que é equivalente a (3.19), pois reconstrói  $x$  utilizando os coeficientes *average* do nível mais alto e os coeficientes *detail* de todos os níveis anteriores.

### 3.3 A Maximum Overlap Discrete Wavelet Transform

Como já mencionado, a Transformada Wavelet tem um grande defeito, a incapacidade de lidar com rotações não pares (i.e. de dois em dois) de um sinal. Isso se traduz por exemplo na mudança dos coeficientes calculados caso uma série temporal (mensal)



se inicie em janeiro ou fevereiro<sup>9</sup>. Percival e Walden [46] apresentam um método que contorna a incapacidade da DWT de lidar com rotações ímpares, baseados na classe dos filtros não decimados<sup>10</sup> chamado de *Maximum Overlap Discrete Wavelet Transform* (MODWT).

Quando aplicamos um banco de filtros a um sinal, podemos ver pelos gráficos das funções ganho quadrado, que nunca são zeradas as frequências exceto aquelas no limite do intervalo, a frequência constante para filtros pa e a frequência de Nyquist para filtros pb. Todas as outras frequências continuam presentes, em amplitudes modificadas. Pelo Teorema da Convolução isso significa que os vetores  $E$  que representam as amostragens dessas frequências são autovetores associados a autovalores não nulos da matriz de convolução que representa o filtro. Agora, se  $H_1$  e  $H_2$  são dois filtros de um banco de filtros e  $Nul_{H_1} \cap Nul_{H_2} = \{\mathbf{0}\}$  então pode-se recuperar  $x$  a partir de  $H_1x$  e  $H_2x$ . Uma prova possível é dada com o uso do Teorema da Convolução.

Seja  $H = (H_1 + H_2)$ , então  $H$  tem DFT dada por  $\mathbf{H}_k = \mathbf{H}_{1k} + \mathbf{H}_{2k}$ , e o fato da interseção dos espaços nulos ter apenas o vetor trivial implica que  $\mathbf{H}_k \neq 0$  para todo  $k$ . Como existe uma bijeção entre os coeficientes de Fourier de um vetor e o próprio vetor, se for realizada a DFT em  $Hx$ , obtendo-se os coeficientes  $\mathbf{H}_k \mathbf{X}_k$ , uma divisão por  $\mathbf{H}_k$ , que supostamente já são conhecidos, retorna o coeficiente da DFT de  $x$  e resta fazer uma aplicação da equação (2.8).

Isso também implica que a imagem das duas matrizes tem interseção não trivial. Portanto existe uma redundância em aplicar os dois filtros. Olhando para os gráficos das funções ganho quadrado do filtro de Haar, fica evidente que para recuperar o sinal, digamos do filtro pb, seria necessário apenas a informação adicional  $\mathbf{X}_{\frac{N}{2}}$ . A heurística da subamostragem é eliminar as redundâncias e gerar uma transformada única, a DWT, que tem posto cheio, seu preço é a perda de informação relativa às rotações.

O primeiro nível da MODWT de Percival e Walden [46] consiste em gerar e aplicar ao sinal as matrizes  $G^1$  e  $W^1$ , mas em vez de usar diretamente  $g$  e  $w$  eles são pré-multiplicados por  $(\sqrt{2}^{-1})$ . A partir de agora todos os filtros, a não ser que dito o contrário,

---

<sup>9</sup>março não faria diferença.

<sup>10</sup>Dizimação é um outro nome dado para subamostragem.

serão ajustados dessa forma. Como já mencionado, isso implica que  $\mathcal{G}(k) + \mathcal{W}(k) = 1$  para todo  $k$ . Temos:

$$G^1 x = \mathbf{a}^1, \mathbf{a}^1 \in \mathbb{R}^{N \times 1} \quad (3.30a)$$

$$W^1 x = \mathbf{d}^1, \mathbf{d}^1 \in \mathbb{R}^{N \times 1} \quad (3.30b)$$

Visto de outra forma, os coeficientes *average* do primeiro nível da MODWT são o vetor  $x_l$  antes de ser subamostrado e os coeficientes *detail* são o vetor  $x_h$ . É direto ver que  $G^1(\mathcal{T}x) = \mathcal{T}\mathbf{a}^1$ , e portanto a transformada é imune a rotações, exceto por uma rotação. Como não é necessária uma subamostragem, não será necessária uma sobre-amostragem então a inversão da operação é facilitada. A matriz de correlação cruzada de  $g$  é  $(G^1)^T$  e de  $w$  é  $(W^1)^T$ . Fazendo

$$(G^1)^T \mathbf{a}^1 = \hat{\mathbf{a}}^1 \quad (3.31a)$$

$$(W^1)^T \mathbf{d}^1 = \hat{\mathbf{d}}^1 \quad (3.31b)$$

pela equação (2.28) temos que a DFT de  $\hat{\mathbf{a}}^1$  é  $\mathcal{G}(k)\mathbf{X}_k$  e a DFT de  $\hat{\mathbf{d}}^1$  é  $\mathcal{W}(k)\mathbf{X}_k$ . Portanto o vetor  $\hat{\mathbf{a}}^1 + \hat{\mathbf{d}}^1$  tem DFT dada por  $(\mathcal{G}(k) + \mathcal{W}(k))\mathbf{X}_k = \mathbf{X}_k$  logo

$$x = \hat{\mathbf{a}}^1 + \hat{\mathbf{d}}^1 = (G^1)^T Gx + (W^1)^T Wx, \quad (3.32)$$

que pode ser escrito no formato de uma MRA

$$x = \mathbf{S}^1 + \mathbf{D}^1.$$

Isso pode ser visto como uma decomposição de  $x$  em duas partes, cada uma correspondendo a um banda de frequências no domínio espectral. O vetor  $\mathbf{S}^1$  é equivalente a filtrar o sinal com um filtro com DFT dada por  $\mathcal{G}$ , com banda  $[0, 1/4]$  e conseqüentemente  $\mathbf{D}^1$  equivale a filtrar o sinal com um filtro de DFT  $\mathcal{W}$ , na banda  $[1/4, 1/2]$ . Juntos eles filtram todo espectro do sinal e é imediato que são filtros com fase zero [46].

Os passos seguintes da transformada vão ser aplicações sucessivas de filtros nos coeficientes *average* dos níveis anteriores. Assim é possível utilizar apenas os resultados já vistos para filtros sem necessidade de se falar nas Transformadas Wavelets para usar a MODWT. Porém muitas coisas úteis que podem ser utilizadas são provadas com a DWT.

Percival e Walden [46] constroem a MODWT de forma que ela tenha uma ligação direta com a DWT e assim várias propriedades desta podem ser utilizadas quando o uso daquela. O primeiro nível da MODWT, após um redimensionamento, é o conjunto  $(x_l, x_h)$  que pode ser subamostrado para gerar o primeiro nível da DWT. Essa relação pode ser descrita de outra forma. Se for realizada a DWT no vetor  $x$  e no vetor  $\mathcal{T}x$  e os coeficientes, após redimensionamento, forem ordenados alternadamente:  $(x_{11}, x_{21}, x_{12}, x_{22}, \dots)$ , será obtida a MODWT, isto é, ela é equivalente a realizar a DWT em  $x$  e em uma rotação de modo a produzir duas amostras de coeficientes.

Assim como cada nível da DWT divide o espectro de um sinal em bandas de frequências, a MODWT também o faz. De fato cada nível  $m$  é calculado aplicando-se as matrizes  $G^m$  e  $W^m$  aos coeficientes *average* do nível anterior, sem necessidade de subamostragem. Assim

$$\begin{aligned}\mathbf{a}^m &= G^m \mathbf{a}^{m-1} \\ \mathbf{d}^m &= W^m \mathbf{a}^{m-1}\end{aligned}\tag{3.33}$$

Sucessivas aplicações do Teorema da Convolução garantem que os vetores em (3.33) têm conteúdo espectral, desconsiderando a fase, respectivamente, dado por [46]:

$$\begin{aligned}\left| \mathcal{G}(2^m k) \prod_{j=1}^{m-1} \mathcal{G}(2^j k) \right| \\ \left| \mathcal{W}(2^m k) \prod_{j=1}^{m-1} \mathcal{G}(2^j k) \right|\end{aligned}\tag{3.34}$$

Para inversão de cada nível, estudamos cada  $\mathbf{a}^m$  individualmente. Pela equação (3.32) trocando  $x$  por  $\mathbf{a}^m$ ,  $G^1$  por  $G^m G^{m-1} \dots G$  e  $W^1$  por  $W^m G^{m-1} \dots G$  temos

$$\begin{aligned}\hat{\mathbf{a}}^{m-1} &= \hat{\mathbf{a}}^m + \hat{\mathbf{d}}^m, \\ \hat{\mathbf{a}}^m &= (G^m G^{m-1} \dots G^1)^T \mathbf{a}^m \\ \hat{\mathbf{d}}^m &= (W^m G^{m-1} \dots G^1)^T \mathbf{d}^m\end{aligned}\tag{3.35}$$

A prova consiste em notar que  $\mathcal{G}(2^m k) + \mathcal{W}(2^m k) = 1$ , já que estas são simplesmente contrações de  $\mathcal{G}$  e  $\mathcal{W}$ , e portanto vale o que foi visto na inversão do primeiro nível [46].

Desta forma é obtida uma decomposição do vetor  $x$  em diversas bandas de frequência. Por (3.35), podemos escolher como representar a equação (3.33):

$$x = \hat{\mathbf{a}}^1 + \hat{\mathbf{d}}^1 = \hat{\mathbf{a}}^2 + \hat{\mathbf{d}}^2 + \hat{\mathbf{d}}^1 = \dots = \hat{\mathbf{a}}^m + \hat{\mathbf{d}}^m + \hat{\mathbf{d}}^{m-1} + \hat{\mathbf{d}}^{m-2} + \dots + \hat{\mathbf{d}}^1$$

em formato de MRA

$$x = \mathbf{S}^1 + \mathbf{D}^1 = \mathbf{S}^2 + \mathbf{D}^2 + \mathbf{D}^1 = \dots = \mathbf{S}^m + \mathbf{D}^m + \mathbf{D}^{m-1} + \dots + \mathbf{D}^1 \quad (3.36)$$

Onde novamente tem-se a identidade  $\mathbf{S}^{m-1} = \mathbf{D}^m + \mathbf{S}^m$ . Cada um dos vetores que decompõe  $x$  em (3.36) é identificado como a parte do sinal composta pelas frequências em  $[0, 1/2^{m+1}]$  para  $\mathbf{S}^m$  e  $[1/2^{m+1}, 1/2^m]$  para  $\mathbf{D}^m$ . Embora essa identificação seja no domínio da frequência, ainda é possível a localização por escalas, já que a MODWT é composta por várias amostras da DWT. De fato, em cada nível  $J$  são realizadas  $2^J$  transformadas wavelets. Já que não são feitas subamostragens, cada nível da MODWT é na verdade a DWT aplicada à diversas rotações dos vetores *average* dos níveis anteriores [46].

Obtém-se assim um método de decomposição baseado em wavelets que é imune a rotações, i.e momento inicial da amostragem do sinal. A desvantagem é o número maior de cálculos necessários, mas para aplicações em estatística, essa desvantagem é mais do que compensada, especialmente por que muitas vezes, devida a baixa quantidade de dados disponíveis, o número de cálculos necessários é pequeno de qualquer forma.

### 3.3.1 Remoção de ruídos com wavelets

Embora a MOWDT seja mais versátil que a DWT, a maioria das propriedades estatísticas que as wavelets possuem são provadas para a DWT. Esta última tem a grande vantagem de ser uma transformada ortogonal. Como já foi mencionado, uma possibilidade de utilização de filtros é a remoção de ruídos de um sinal e por ser uma transformada ortogonal, a DWT mantém a média e a variância amostral de um sinal, e se o erro for aditivo e a wavelet bem escolhida, ela separa a variância do sinal da variância do ruído, pelo menos aproximadamente. Além disso, ela tem outra vantagem, pois como foi mencionado na Seção 3 ela é capaz de decompor vetores em  $\mathbb{R}^N$  em uma base esparsa. Uma instrumentalização destas propriedades é chamada *thresholding* e para o caso das wavelets foi mostrada por Donoho e Johnstone [17].

Percival e Walden [46] e Walker [58] fazem a seguinte observação: suponha que  $\mathcal{O}$  é uma matriz ortogonal e que  $c = \mathcal{O}x$ . Temos que  $\|c\|^2 = \|x\|^2$ , e, pelo Teorema de Parseval, escrevendo  $c = (c_0, c_1, \dots, c_{N-1})$

$$\sum_{i=0}^{N-1} c_i^2 = \sum_{i=0}^{N-1} x_i^2.$$

A remoção de ruídos de um sinal é uma maneira diminuir sua energia, no sentido de que encontra-se um sinal próximo de  $x$  com energia (norma) menor, mas que mantenha a maior parte do que chama-se tendência do sinal. Em um gráfico isso equivale a suavizar a curva resultante da interpolação dos pontos. Para mostrar o caso geral, Percival e Walden [46] supõem que  $c'$  seja uma ordenação crescente (em módulo) de  $c$ , isto é,

$$|c'_0| \geq |c'_1| \geq \dots \geq |c'_{N-1}|.$$

A função perda de não se usar todos os coeficientes de  $c'$  para reconstruir  $x$  é dada por:

$$loss_c(K) = \frac{\sum_{i=0}^K |c'_i|^2}{\|c\|^2}. \quad (3.37)$$

A função  $loss$  retorna, em percentual, e para cada  $K$ , quanta informação foi removida ao reconstruir-se  $x$  com os coeficientes selecionados. Escolhe-se um nível de confiança,  $(1 - \varepsilon)\%$ , e.g.  $\varepsilon = 0.05$ ,  $loss_c(K) = 0.95$ . Isso equivale a encontrar uma matriz  $Q$  tal que

$$\|x - \mathcal{O}^T Q c\| \leq \varepsilon. \quad (3.38)$$

A matriz  $Q$ , chamada de projeção diagonal, é uma matriz diagonal onde as entradas diferentes de 0 são aquelas que foram escolhidas com a função  $loss$ , isto é, se  $c'_k$  não foi utilizado, então encontra-se o  $c_j$  correspondente e  $[Q]_{j,j} = 0$ .

Uma generalização de  $Q$  é uma matriz  $Q_\delta$  que também é diagonal mas em vez de ter entradas 0 ou 1 tem entradas 0 ou  $\delta$ . A questão é saber se é possível escolher um específico  $\delta$  que melhore a estimativa do que é sinal do e do que é ruído. O resto desta seção vai apresentar o caso quando a projeção ortogonal usada é a DWT.

Para utilizar a DWT em aplicações estatísticas, supomos que são feitas  $n = k2^J$  observações de um sinal, onde  $k$  e  $J \in \mathbb{N}$  e o espaço de tempo entre duas observações é fixo para todo  $n$ . A amostra precisa ser um múltiplo de  $2^J$  para ser possível aplicar

a DWT  $J$  vezes. Obtém-se desta forma um vetor  $y = (y_0, y_1, y_3, \dots, y_{N-1})$ . Mas agora em vez deste vetor ser o objeto de interesse, ele é na verdade composto por uma função desconhecida, a qual gostaríamos de estimar mais um termo de ruído:

$$y_t = x_t + \epsilon_t \quad (3.39)$$

onde  $x = f(t)$ . Esta função pode ser determinística ou estocástica. Donoho e Johnstone [17] desenvolvem o método para funções determinísticas, enquanto outros autores (ver capítulo 10 de Percival e Walden [46]) desenvolvem para funções estocásticas. Neste trabalho seguiremos o primeiro, isto é,  $f(t)$  é uma função determinística e  $\epsilon_t$  é uma variável aleatória tal que  $\epsilon \sim \mathcal{N}(0, \sigma^2)$ .

Assim como os sinais que foram discutidos anteriormente, assume-se que  $f$  está definida no intervalo  $[0, 1]$ , e que  $t_i = \frac{k}{N}$ . para  $k = 0, 1, 2, \dots, N-1$ , isto é o vetor  $x$  é uma amostra de tamanho  $N$  de uma função desconhecida. Donoho and Johnstone [17] sugerem encontrar um vetor  $\hat{f} = (\hat{f}_i)_{i=0}^{N-1}$  que estime o vetor  $f = (f)_{i=0}^{N-1}$ , e na norma euclidiana minimize o risco:

$$R(\hat{f}, f) = n^{-1} E \|\hat{f} - f\|_2^2 \quad (3.40)$$

Minimizar a equação (3.40) equivale a encontrar o vetor  $\hat{f}$  de menor distância (na norma euclidiana) possível para esta  $f$  desconhecida.

Donoho and Johnstone [17] consideram métodos espaciais adaptativos, ou simplesmente métodos adaptativos, para estimar  $\hat{f}$ . Um método adaptativo é uma forma de estimação que não impõe ao estimando ser uma única função em todo intervalo. A regressão linear clássica é um exemplo de método não adaptativo, onde um vetor qualquer é projetado linearmente sobre o espaço gerado por um grupo de funções (que podem ser lineares ou não) e obtemos a melhor aproximação possível, minimizando os erros quadrados entre  $y$  e sua projeção  $\hat{y}$ . Um método adaptativo continua fazendo projeções, mas ao invés de minimizar a distância entre  $y$  e uma única função do espaço, divide o intervalo de definição de  $y$  em sub intervalos, digamos  $I_j$ , disjuntos, e procura qual a função do espaço que é mais próxima de  $y$  em cada sub intervalo. Donoho and Johnstone [17] dão como exemplo de método adaptativo a aproximação  $y$  por polinômios de grau  $D$ . O intervalo  $[0, 1]$  é dividido em  $K$  sub intervalos do tipo  $I_j = [\delta_j, \delta_{j+1}]$ , onde cada  $\delta_j$

pertence a coleção  $\Delta = \{0, \delta_1, \delta_2, \dots, \delta_K\}$  e  $p_j(t)$  é o polinômio que minimiza:

$$\|y - p_j(t)\|^2, y \in I_j$$

$$\hat{f} = \sum_{k \in \Delta} p_j(t) 1_{I_j}$$

Onde a função 1 é o indicador do intervalo.

Um método adaptativo ideal é aquele que retorna o menor risco medido por (3.40), isto é, a escolha dos intervalos  $I_j$  que minimiza  $R$  ao projetar-se  $y$  sobre o espaço dos polinômios de grau até  $D$ . Em geral esse método não existe, mas Donoho and Johnstone [17] supõem que exista um oráculo que possa indicar qual a partição ótima de  $[0, 1]$  que minimiza  $R$ . Este oráculo retornaria uma escolha de  $\Delta$  que minimiza (3.40). Este risco ideal, para uma projeção ortogonal, é dado por  $N^{-1} \times (\text{número de parâmetros estimados}) \times (\text{variância do ruído})$ . Conhecemos o número de parâmetros estimados, que é  $K \times (D + 1)$  e também a variância do ruído  $\sigma^2$  e assim obtém-se o risco ideal:

$$\frac{K \cdot (D + 1) \cdot \sigma^2}{n} = \frac{\#parâmetros \cdot \sigma^2}{n} \quad (3.41)$$

Não se espera atingir (3.41) em nenhum momento quando se lida com dados empíricos, mas sendo o menor risco que pode ser esperado, o risco ideal serve para medir a qualidade de seleção de  $\Delta$  em um método adaptativo, ou seja, a melhor escolha de subintervalos para estimar  $\hat{f}$  é aquela que mais aproxima (3.40), o risco efetivo, de (3.41), o risco ideal.

Nota-se que ao proceder desta forma, não estima-se a função  $f$ , mas os seus valores. Após aplicar o algoritmo piramidal de Mallat [43] nos dados espera-se que seja possível separar os valores da transformada que se devem ao sinal e aos que se devem ao ruído. Se isso for possível, ao reconstruir a série, teremos um conjunto de valores que foram gerados (com alguma certeza) apenas pela função  $f$ .

Seja  $P$  a matriz que representa a primeira etapa da DWT apresentada na seção 3.1, para qualquer nível  $J$  desejado, isto é,  $Px$  gera  $\mathbf{a} = \mathbf{a}^J$  e  $\mathbf{d} = [\mathbf{d}^J, \mathbf{d}^{J-1}, \dots, \mathbf{d}^1]$ . Suponha que eles tenham sido obtidos e ordenados em um único vetor  $\mathbf{c}$  da mesma forma que em [58], isto é  $\mathbf{c} = [\mathbf{a}, \mathbf{d}]$  é o vetor dos coeficientes da DWT. Já que  $P$  é uma transformada linear ortogonal,  $y = P^T \mathbf{c}$ . Dependendo da wavelet escolhida e da função  $f$ , os vetores  $\mathbf{d}^1, \mathbf{d}^2, \dots, \mathbf{d}^{J-k}$ , onde  $0 \leq k < J$ , serão mais ou menos esparsos.

Johnstone [38] chama de esparsidade a propriedade de algum vetor ter seus valores de maior magnitude concentrados em algumas posições. Por exemplo o vetor  $(1,0)$  é mais esparsos que o vetor  $(1/\sqrt{2}, 1/\sqrt{2})$ . Lembrando que a energia de um vetor é dada pela norma 2 e segue que os dois vetores do exemplo tem a mesma energia, 1. Se forem tomada outra norma, digamos a norma  $p$  então a norma do primeiro vetor será maior do que a norma do segundo vetor.

Como  $P$  é uma isometria, ela preserva a norma euclidiana do vetor  $y$ , mas ela aloca a energia em cada nível de forma desigual entre  $\mathbf{a}$  e  $\mathbf{d}$ , no caso, a norma dos vetores *average* tende a ser maior do que a dos vetores *detail* [58]. Esta alocação depende da wavelet escolhida e da função  $f$  que compõe o sinal. A heurística do artigo de Donoho e Johnstone [17] é que se a wavelet for escolhida adequadamente, então abaixo<sup>11</sup> de algum nível  $j^*$  os vetores  $\mathbf{d}^{j^*}$  serão convenientemente esparsos ou seja, terão muitas entradas de baixa magnitude e as entradas com grande magnitude serão poucas e espalhadas. Como  $y = \sum_{i=0}^{N-1} c_i P^T(i)$  onde  $c_i$  é a  $i$ -ésima entrada de  $\mathbf{c}$  e  $P^T(i)$  é a  $i$ -ésima coluna de  $P^T$ , se algum vetor  $d_j$  for esparsos, apenas algumas entradas de  $\mathbf{c}$  nesse nível serão significativas e supostamente pode-se remover as insignificantes. O que é preciso saber é quais delas podem ser removidas e como isso vai afetar o risco (3.40) medido.

Donoho e Johnstone [17] fixam uma wavelet que tenha número de *vanishing moments*  $M$  para resoluções menores que  $j^*$  e suporte  $S$ . Isso quer dizer que se  $f$  for aproximadamente um polinômio de grau até  $M$  no intervalo de suporte de uma das wavelets em um nível  $j \leq j^*$  então o coeficiente *detail* neste nível e intervalo será zero. Por outro lado se  $f$  tiver uma mudança estrutural no intervalo,  $\mathbf{d}$  irá captar esse movimento. Outro fato importante é que como  $P$  é linear e ortogonal:

$$Py = Pf + P\epsilon = \mathbf{c}_f + \mathbf{c}_\epsilon = [\mathbf{a}_f, \mathbf{d}_f] + [\mathbf{a}_\epsilon, \mathbf{d}_\epsilon] \quad (3.42)$$

$$\mathbf{d}_\epsilon \sim \mathcal{N}(0, \sigma^2)$$

Isto é, os coeficientes obtidos ao aplicar  $P$  em  $y$  são a soma dos coeficientes obtidos ao aplicar-se  $P$  em  $f$  e em  $\epsilon$ . Estes últimos preservam a distribuição original do

---

<sup>11</sup>Neste trabalho estamos indexando os coeficientes da DWT de forma crescente, i.e. os níveis mais baixos usam intervalos menores. Donoho e Johnstone [17] indexam os coeficientes da DWT de forma decrescente, i.e. níveis mais altos utilizam intervalos menores.



ruído [17]. Chama-se os coeficientes de  $\mathbf{c} = Py$  de coeficientes empíricos e deseja-se extrair  $\mathbf{c}_f$  deles. Aqui entra a ideia de esparsidade para recuperar  $f$ . Suponha que  $f$  de fato seja aproximada por um polinômio de grau até  $D \leq M$  nos intervalos diádicos  $[k/2^{j+1}, k/2^j]$  para todo  $j \leq j^*$ , exceto em um número pequeno de intervalos. Como estamos calculando os coeficientes da expansão em wavelets exatamente nestes intervalos, e devido ao fato da wavelet utilizada ter  $M$  *vanishing moments*, os coeficientes *detail* devidos a  $f$  devem ser zero exceto em um número pequeno de intervalos onde  $f$  tem quebra estrutural, ou seja o vetor  $d_j$  de  $f$  vai ser esparço.

Donoho e Johnstone [17] descrevem um método adaptativo que tira vantagem da propriedade descrita acima. Os autores propõe que seja criada uma lista  $\delta$  de elementos de  $\mathbf{d} = \mathbf{d}_f + \mathbf{d}_\epsilon$  e estima-se  $f$  como:

$$\hat{f} = \sum_{i \in \delta} \mathbf{c}_i P(i) \quad (3.43)$$

Ao estimar  $f$  com (3.43) estamos utilizando apenas alguns coeficientes. Isso, como já foi dito, equivale a encontrar uma matriz diagonal  $Q$  com entradas 0 e  $\delta$ . Pergunta-se qual deve ser a lista  $\delta$  a ser utilizada? A resposta de Donoho e Johnstone[17] é todo  $i$  para os qual os coeficientes  $d_{f_i}$  são diferentes de zero. Se um oráculo nos dissesse qual deve ser a lista a ser tomada então (3.41) se transformaria em:

$$\frac{\#\delta * \sigma^2}{n} = \mathcal{R}_\psi \quad (3.44)$$

Onde  $\mathcal{R}_\psi$  denota o risco ideal de se estimar  $f$  utilizando uma wavelet  $\psi$ . Observa-se que ao se mudar a wavelet o risco não será o mesmo. O oráculo apenas fornece o menor risco ao se utilizar uma função estimadora fixa, não o menor risco dentre todas as funções estimadoras possíveis. Tendo este limite inferior para o risco, Donoho e Johnstone [17] estudam quão próximo podemos esperar chegar dele utilizando algum método de thresholding para gerar  $\delta$ , ou seja qual o limite superior para o risco. A resposta é que utilizando o método descrito abaixo, os autores provam que é possível limitar o risco (3.40), desde que se conheça  $\sigma^2$ , ao estimar com wavelets, para:

$$R(\hat{f}, f) \leq (2 \log(n) + 1) \left\{ \mathcal{R}_\psi + \frac{\sigma^2}{n} \right\} \quad (3.45)$$

### 3.3.1.1 Adaptação espacial com soft Thresholding.

Seja  $\mathbf{c} = Py$ . Suponha que:

$$\mathbf{c}_i = \theta_i + \sigma z_i \quad (3.46)$$

Onde queremos estimar  $\theta$ , os coeficientes  $\mathbf{c}_f$ , por algum  $\hat{\theta}$ . O vetor  $z$  é uma realização de uma variável aleatória com distribuição normal e média 1 e  $\sigma$  é o nível do ruído. Donoho e Johnstone [17] definem o risco de estimação como:

$$R(\hat{\theta}, \theta) = E\|\hat{\theta} - \theta\|_2^2 \quad (3.47)$$

Utilizando projeção diagonal  $Q_\delta$  multiplica cada coordenada de  $\mathbf{z}$  por um  $\delta_j$ , e uma projeção ideal fornecida por um oráculo iria manter apenas os coeficientes  $z_i$  que são maiores que o nível do barulho, ou seja  $\delta_j = 0$ , se  $z_j \leq \sigma$ , fornecendo o risco ideal [17]:

$$\mathcal{R}_{Q_\delta} = \sum_{i=1}^n \min(|\theta_i|^2, \sigma^2). \quad (3.48)$$

Donoho e Johnstone [17] definem *soft threshold* como:

$$\eta_S(x, \lambda) = \text{sign}(x)(|x| - \lambda)_+ \quad (3.49)$$

Onde a função  $\text{sign}(x)$  é o sinal do número  $x$  e  $(\cdot)_+^{12}$  é a parte positiva de  $|x| - \lambda$ . Aplicando (3.49) em cada coeficiente de  $\mathbf{c}$  obtemos uma projeção diagonal  $Q_\lambda$  que elimina todos os elementos menores do que  $\lambda$  em magnitude e diminui em  $\lambda$  os elementos maiores, para um nível de *threshold*  $\lambda$ . O Teorema 1 em [17] garante que:

$$\hat{\theta}_i^u = \eta_S(\mathbf{c}_i, \sigma(2\log(n))^{\frac{1}{2}})$$

satisfaz

$$R(\hat{\theta}^u, \theta) \leq (2\log(n+1)) \{ \sigma^2 + \mathcal{R}_{Q_\delta} \} \quad (3.50)$$

---

<sup>12</sup> $(x)_+ = \begin{cases} x, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0 \end{cases}$

Para o nível universal de *threshold*  $\lambda = \sigma(2\log(n))^{\frac{1}{2}}$ . Passa-se então a estimação de  $f$  a partir dos dados. O Teorema de Parseval garante que:

$$\|f - \hat{f}\|_2 = \|\theta - \hat{\theta}\|_2$$

Assumindo que os dados seguem o modelo (3.39) e que foram obtidos os coeficientes empíricos (3.46). Então o risco (3.40) é dado por [17]:

$$\begin{aligned} n^{-1}E\|f - \hat{f}\|_2 &= n^{-1}E\|\theta - \hat{\theta}\|_2 \\ &\leq (2\log(n+1)) \left\{ \frac{\sigma^2}{n} + \mathcal{R}_\psi \right\} \end{aligned} \quad (3.51)$$

Onde  $f$  é a função a ser estimada e  $\theta$  são seus coeficientes verdadeiros,  $\hat{f}$  é uma função que tem sua expansão em wavelets dada por  $P^T Q_\lambda \mathbf{c}$ , ou seja, uma função construída a partir dos coeficientes empíricos nos quais foi realizado o processo de *thresholding*. Dessa forma descobre-se uma banda na qual sabe-se com alguma certeza que os valores de  $f$  estão, ou seja, foram separados, aproximadamente, os valores de  $f$  e do ruído.

A extensão do algoritmo de *thresholding* para a MODWT foi dada por Percival e Walden [46]. Ela consiste em notar que a MODWT é uma DWT aplicada a rotações do vetor  $x$  com os coeficientes intercalados. Para  $N$  múltiplo de 2, o primeiro nível da MODWT é equivalente a aplicar a DWT em  $x$  e  $\mathcal{T}x$  e intercalar os coeficientes, após um redimensionamento, isto é, se  $P$  é a matriz que define a DWT para o primeiro nível e se  $\mathbf{a}_x^1$  são os coeficientes *average* de  $Px$  e  $\mathbf{a}_{\mathcal{T}^{-1}x}^1$  são os coeficientes *average* de  $P(\mathcal{T}^{-1}x)$  então se  $M$  é a matriz que define a MODWT para o primeiro nível, os coeficientes *average* ( $\mathbf{a}^1$ ) de  $Mx$  são dados por:

$$\mathbf{a}^1 = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{a}_{\mathcal{T}^{-1}x,1}^1 \\ \mathbf{a}_{x,1}^1 \\ \mathbf{a}_{\mathcal{T}^{-1}x,2}^1 \\ \mathbf{a}_{x,2}^1 \\ \vdots \\ \mathbf{a}_{\mathcal{T}^{-1}x,N/2}^1 \\ \mathbf{a}_{x,N/2}^1 \end{bmatrix}. \quad (3.52)$$

O mesmo acontece com os coeficientes *detail*. Para os níveis superiores o padrão se mantém, a MODWT equivale a aplicar a DWT a vetores *average* e *detail* do nível anterior, rotacionados. Por outro lado, também podemos ver a DWT como a MODWT

com subamostragem. Nesse caso não nos preocupamos com o tamanho de  $N$  e vemos os coeficientes como uma amostra maior de coeficientes obtidos pela DWT.

Percival e Walden [46] ajustam o nível de *threshold* para ele ficar compatível com os coeficientes da MODWT. Novamente, com  $\lambda = \sigma(2\log(n))^{\frac{1}{2}}$ , como a MODWT redimensiona os valores das entradas de  $P$  por um fator de  $1/2^j$  onde  $j$  é o nível da transformada, o nível de *threshold* também deve ser redimensionado dependendo do nível no qual ele está sendo aplicado, pelo mesmo fator. Assim o *threshold* da MODWT é dado por:

$$\lambda = \sigma \left( \frac{2\log(n)}{2^j} \right)^{\frac{1}{2}} \quad (3.53)$$

O resto do processo é idêntico. Seleciona-se os coeficientes como antes, isto é, formando a matriz diagonal  $T_\lambda$ , e lembrando que as matrizes que formam a MODWT são  $G$  e  $W$ , fazemos:

$$\hat{f} = G^T \mathbf{a} + W^T Q_\lambda \mathbf{d}, \quad (3.54)$$

que é a MODWT após sofrer um processo de *threshold*.

O último detalhe necessário para aplicar o processo de *threshold* nos coeficientes empíricos é conhecer  $\sigma$ . Em geral isso não é possível, portanto ele deve ser estimado. Donoho e Johnstone [17] sugerem utilizar o desvio absoluto mediano ( $\hat{\sigma}_{mad}$ ) dos coeficientes do primeiro nível *detail* dividido por 0.6745. Está divisão se dá para que  $\hat{\sigma}_{mad}$  seja um estimador para o desvio padrão da distribuição Normal [46]. Neste trabalho será usado este estimador de  $\sigma$ , mas redimensionado para a MODWT, segundo Percival e Walden [46]:

$$\hat{\sigma} = \frac{2^{\frac{1}{2}} \text{mediana}(|\mathbf{d}_0^1|, |\mathbf{d}_1^1|, |\mathbf{d}_2^1|, \dots, |\mathbf{d}_{N-1}^1|)}{0.6745} = \frac{2^{\frac{1}{2}} \hat{\sigma}_{mad}}{0.6745}. \quad (3.55)$$

Onde a multiplicação por  $2^{\frac{1}{2}}$  é por causa do redimensionamento dos coeficientes da MODWT. Como  $E(\mathbf{d}_j^1) = 0$  os desvios são os próprios valores de  $\mathbf{d}_j^1$ .

## 4 MEDINDO O NÚCLEO DE INFLAÇÃO

### 4.1 Por que o núcleo de inflação.

As notícias do começo do ano de 2018 no noticiário financeiro internacional foram (volta da) volatilidade e de preocupação com a inflação nos Estados Unidos. Um aquecimento da economia faz pressão para os preços americanos subirem e em geral quando isso ocorre os agentes econômicos esperam subidas na taxa de juros básica da economia, isto é, a taxa usada para política monetária pelo Federal Reserve<sup>1</sup> (FED). O Banco Central do Brasil (BACEN ou BC) faz menção a este fato em seu próprio relatório da inflação [6], na parte onde analisa a conjuntura externa. Em março de 2018 a perspectiva era de um possível aperto monetário nas economias avançadas para evitar o crescimento da inflação. E o principal canal pelo qual a autoridade monetária efetiva um aperto monetário é através da taxa de juros de títulos públicos. Segundo o BACEN [6] ele deve afetar a volatilidade financeira global e afetar a precificação dos ativos em países emergentes.

Nos Estados Unidos, a taxa básica de juros da economia é dada pelos *Federal Funds*, que são títulos de curtíssimo prazo (um dia), negociados entre instituições financeiras deficitárias e superavitárias por meio do Federal Reserve. Os dados podem ser encontrados no site do Federal Reserve of St. Louis [25]. Olhar para as decisões sobre ela ajuda a entender o que o banco central estava pensando naquele momento, especialmente, se ele achava que era necessário intervir na inflação. Porém uma análise completa destas reuniões seria tema para outro trabalho. Consideramos no que segue, a forma mais geral de analisar esta questão, baseada em Blinder [8].

A Figura 4.1 mostra o gráfico do tempo de duração de cada meta para a taxa dos *federal funds* desde janeiro de 2009 até 16 de abril de 2018, em ordem cronológica, isto é, quanto tempo levou para o FED perceber necessidade de mudar a taxa de juro. Ele mostra no eixo das abscissas a meta e no eixo das ordenadas o quanto tempo ela durou. O valor da meta não importa aqui e sim quanto tempo ela ficou em vigência e quantas vezes ela foi modificada. O fato de a taxa de juros ter ficado por 2500 dias consecutivos

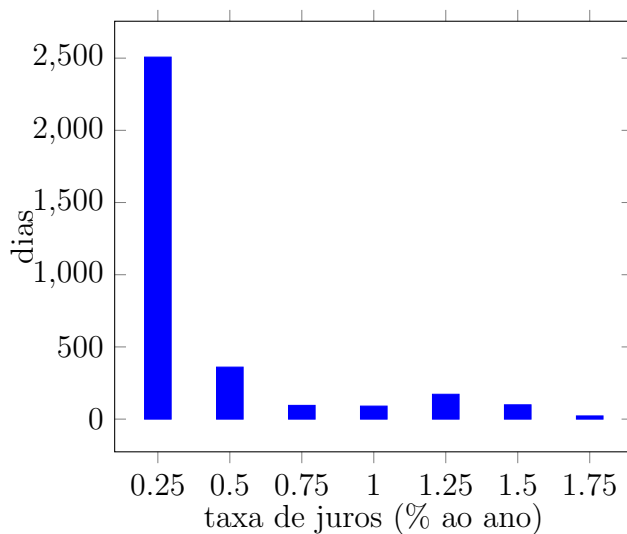
---

<sup>1</sup>Banco central dos Estados Unidos

no nível de 0.25% ao ano e a partir de dezembro de 2015 ter começado a subir, com as decisões de aumento sendo tomadas em intervalos menores que 2500 dias, mostra que alguma mudança na economia foi percebida pelo FED que necessitou uma intervenção na política monetária. Essas mudanças podem ter motivos diversos, mas as razões de se controlar a taxa de juros é a inflação e o emprego, como exposto em uma minuta do Federal Reserve:

Ao decidir a política monetária, o comitê procura mitigar os desvios da inflação de seu alvo de longo prazo e desvios do emprego de seu nível máximo estimado pelo comitê. Federal Reserve of The United States [26, p. 8, tradução do autor].

**Figura 4.1:** Tempo de vigência em dias da taxa meta dos *Federal Funds*



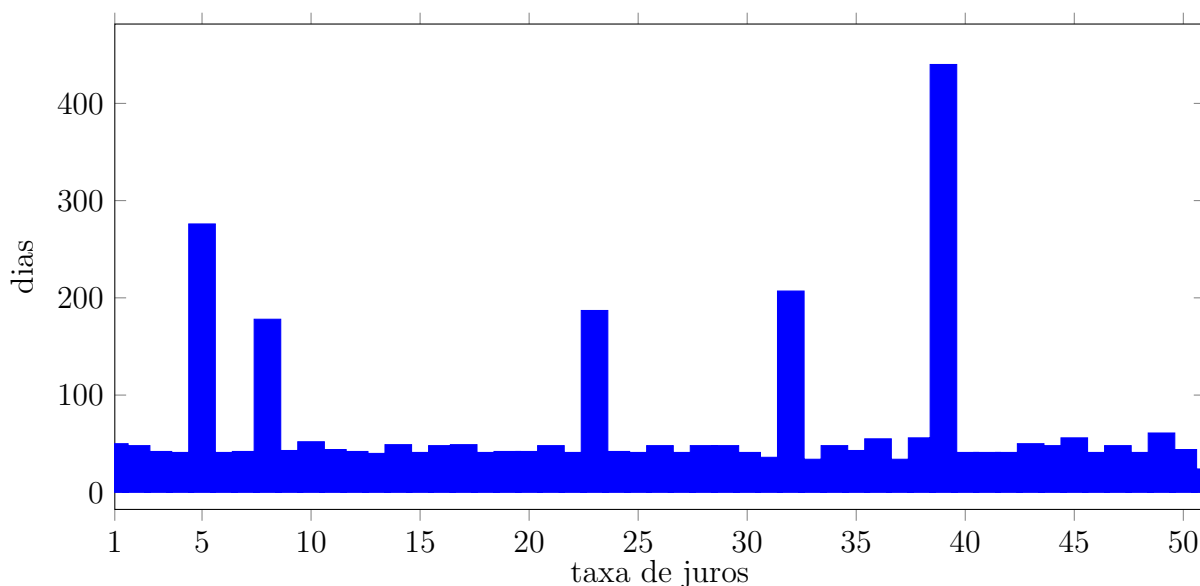
Fonte: Federal Reserve of St. Louis [25].

Notamos que a relação entre inflação e emprego é conhecida como curva de Phillips, que diz que um aumento (redução) na inflação, ocasiona temporariamente uma redução (aumento) na taxa de desemprego.

Contrastamos com a Figura 4.2, das mudanças da taxas de juro no Brasil, a SELIC meta, para o mesmo período. Novamente as barras indicam quanto tempo uma determinada meta durou até a próxima mudança. Nesta figura são omitidas os valores das taxas de juro para não comprometer a clareza do gráfico (existem 29 metas diferentes e a

ordem crescente/decrescente delas não é cronológica), os números no eixo das abscissas são a posição cronológica de cada meta. Foram feitas neste período de 9 anos, 49 alterações na taxa de juros brasileira, contra 3 nos Estados Unidos. Isto é um argumento para mostrar a diferença na situação monetária nos dois países durante os anos mais recentes. Os Estados Unidos tiveram uma taxa de inflação controlada, o que requiriu de seu banco central menos intervenções. O Brasil por outro lado teve diversos períodos e mudanças na política econômica em geral, que foram seguidas pela política monetária. Formalmente, em seus objetivos, não em suas decisões, esta última continuou a mesma em todo período, pois continuou-se a perseguir a mesma meta de inflação.

**Figura 4.2:** Tempo de vigência em dias da taxa meta SELIC

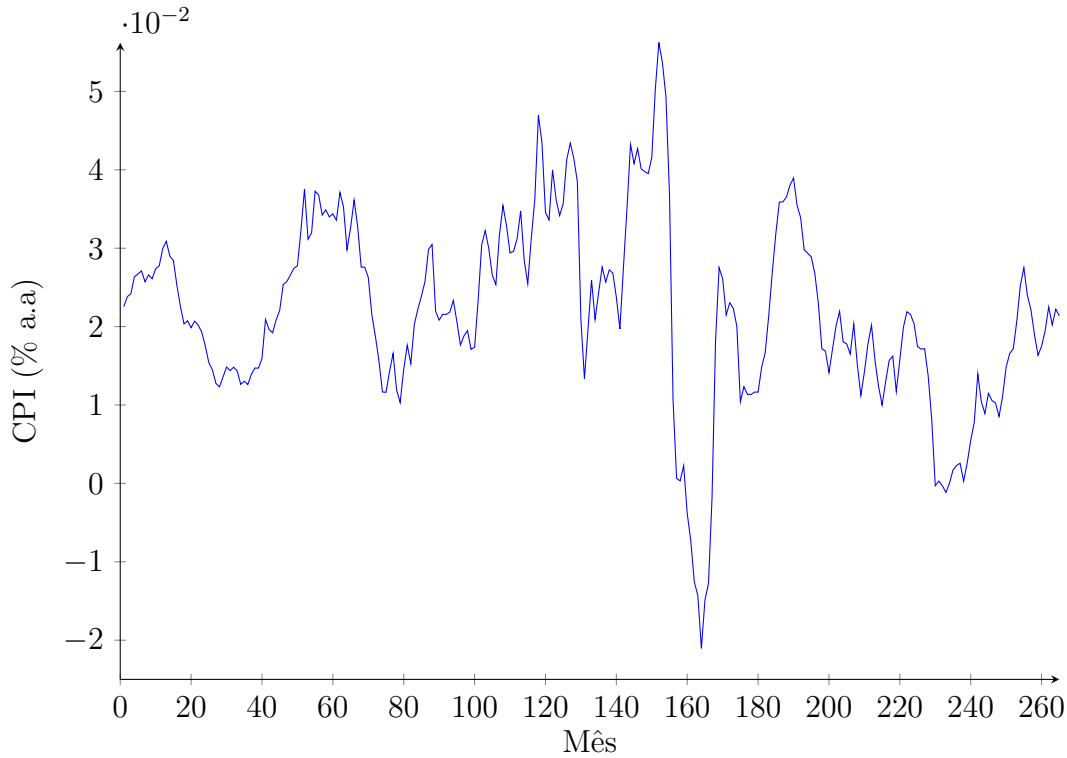


Fonte: Banco Central do Brasil [3].

A figura 4.3 mostra a taxa de inflação americana medida pelo Consumer Price Index (CPI) de dezembro/1995 até dezembro 2017. Talvez no gráfico não fique claro por que o FED tomou apenas três decisões de alteração em 9 anos. A parte do gráfico que reporta este período são as 108 últimas observações, por volta da observação de número 160. Mas olhemos para outras métricas. No período de 9 anos entre janeiro/2009 e dezembro/2017 a taxa de inflação ficou abaixo da meta do FED de 2% ao ano [26] por 65% do tempo, abaixo de 2.5%a.a. por 84% do tempo e abaixo de 3%a.a. por 92% do tempo. Ela foi negativa em 10% das observações. Comparando com o IPCA brasileiro (Figura 4.4 na seção 3.2) no mesmo período, a taxa de inflação ficou acima da meta do

Banco Central de 4.5%a.a. ao ano por 85% do período e acima dos 2%a.a. de tolerância por 37.9% do período. Isso mostra por que foi necessário menos alterações na taxa de juros nos Estados Unidos.

**Figura 4.3:** CPI mensal à taxas anualizadas entre dez/1995 e dez/2017



Fonte: Bureau of Labor Statistics [11]

Um aspecto da tomada de decisões dos bancos centrais é o gradualismo com o qual elas são tomadas. Blinder [8] menciona que bancos centrais têm tendência a serem graduais na mudança de política monetária, e por gradual ele assume que as mudanças de metas não sejam maiores que 0.25% para cima ou para baixo. Embora não dê uma resposta definitiva do porquê o gradualismo, ele aponta cinco possíveis motivos [8, p. 36]: (i) valor da opção, pois mudanças graduais erradas têm efeitos adversos menores que mudanças bruscas e a autoridade monetária tem a opção de corrigir a decisão; (ii) choques serialmente correlacionados, o que impede a verificação tempestiva do efeito de uma política; (iii) conservadorismo breinardiano (de William Breinard) que considera que dependendo do tipo de incerteza com a qual o formulador de política se depara, ele vai tomar medidas diferentes; (iv) preocupação com a volatilidade dos mercados, pois o banco central pode considerar que aumentos bruscos podem aumentar em demasia a volatilidade; e por



último, (v)por que ajustes graduais no curto prazo podem levar a reações fortes nos juros de longo prazo.

Seja qual for a razão, o gradualismo foi efetivamente adotado para os Estados Unidos neste período, já que todas mudanças foram de 0.25%. No Brasil, 12 das 49 mudanças foram de 0.25%, enquanto que a média das mudanças, i.e. a media dos valores absolutos das diferenças entre duas metas consecutivas, foi de 0.55%. A maior mudança observada foi de 1%. O tempo médio de duração de uma meta foi de 65 dias. É evidente que, sem verificar as condições econômicas, não é possível determinar o motivo de cada uma das mudanças, nem é o objetivo deste trabalho fazer isso. Por outro lado pode-se perguntar se o motivo que ocasionou as mudanças afeta nossa capacidade de modelar a inflação.

Mas é verdade que em cada uma destas mudanças houve algum acontecimento que convenceu os membros do Comitê de Política Monetária (COPOM), que é o órgão responsável por estabelecer a meta da taxa de juro, de que era oportuno modificar a taxa de juro naquele instante, muitas vezes de forma brusca. Neste contexto uma das ferramentas que bancos centrais podem usar para avaliar a situação, inclusive sendo mencionada pelo BACEN em seus relatórios, é a medida de núcleo de inflação. Não podemos dizer qual a importância dada ao núcleo, mas esta é uma ferramenta disponível. Uma questão que se coloca naturalmente é a seguinte: medidas de núcleo, especialmente as baseadas em wavelets, podem ajudar a tomar melhores decisões?

#### **4.1.1 Núcleo de inflação.**

Segundo Baqaee [2], embora não exista uma definição específica do que seja núcleo de inflação, em geral é aceito que uma boa medida de núcleo deve satisfazer dois atributos: reduzir a volatilidade da inflação medida e ser um bom preditor de inflação futura.

Seguindo a ideia de que o núcleo deve satisfazer o primeiro atributo, Bryan e Checchetti [10, p. 197, tradução do autor] propõem que o núcleo de inflação é o "[...] componente de longo prazo, ou persistente do índice de preços medido[...]". Um banco central (BC), ao assumir a missão de estabilização de preços em uma economia, deve ter meios de medir o efeito das políticas estabilizadores, ou se a variação do nível de preços

observado requer medidas de estabilização. A inflação observada sofre de pelo menos dois efeitos que deveriam ser removidos antes de uma ação estabilizadora do BC [13]: viés e ruídos. O viés pode acontecer por várias razões. Por exemplo, visto que a inflação mede o custo de uma cesta de bens e serviços que os indivíduos em uma economia consomem, se a cesta for uma representação ruim do padrão de consumo, a inflação será uma medida ruim do aumento do custo de vida. Ruídos, por outro lado, são perturbações que acontecem aleatoriamente nos preços de alguns produtos, e em geral esperam-se ser transitórios.

Já Blinder [7] em um comentário sobre o artigo de Checchetti [13] aponta que o núcleo deve ter a capacidade de prever inflação futura. Quando Bryan e Checchetti [10] e Checchetti [13] sugerem que a parte mais importante da inflação é aquela que é expurgada de efeitos transitórios, restando somente uma tendência adjacente, eles implicam que é suficiente saber qual o núcleo de inflação naquele momento para a tomada de decisões pelo Banco Central.

Seja como for, as medidas de núcleo em geral são feitas partindo-se do princípio de remover ruídos transitórios do índice de preços observados. Existem algumas formas bastante difundidas de se fazer isso. No Brasil, o BACEN calcula e apresenta medidas de núcleo de inflação desde o ano de 1999, juntamente com a adoção do plano de metas para a inflação [54]. Entre as cinco medidas publicadas destacam-se a medida por exclusão (EX2), a por médias aparadas (MA) e a por dupla ponderação (DP). Segundo o BACEN, o plano de metas visa ancorar as expectativas dos agentes econômicos quanto à inflação medida pelo Índice de Preços ao Consumidor Amplo (IPCA). Para isso o banco utiliza medidas de núcleo como auxiliares na tomada de decisões, embora o alvo dessas decisões sempre seja o IPCA [3]. Cabe notar que a importância das medidas de núcleo para a política monetária muda de país para país [3, 63]. Em se tratando de banqueiros centrais, nos Estados Unidos, Blinder [8] é a favor de usar a variação do núcleo *ex energia e alimentos* como meta.

A seguir são descritas as medidas de núcleo de inflação calculadas pelo BACEN que serão avaliadas. Elas podem ser classificadas em duas categorias distintas: medidas por exclusão e medidas estatísticas. As primeiras são construídas removendo-se certos itens do cálculo da inflação, sob o argumento de que a oscilação destes itens seja em geral transitória e por isso não deveria ser levada em conta pela autoridade monetária. As

segundas são construídas baseadas em métodos estatísticos de remoção de componentes de grande oscilação. Embora a ideia seja a mesma das medidas por exclusão, as medidas estatísticas não removem sempre os mesmos componentes, mas escolhem quais deles remover, período a período, baseado em alguma relação quantitativa.

**Núcleo por exclusão 2 (EX2).** Segundo o Bureau of Labor Statistics [12] o núcleo por exclusão é calculado por acreditar-se que o CPI<sup>2</sup> expurgado de comida e energia é uma medida melhor do núcleo da inflação, devido à alta volatilidade daqueles componentes. Originalmente a medida por exclusão calculada, agora chamada de núcleo por exclusão 0 (EX0), foi apresentada no relatório da inflação de março de 2001 e consistia em remover os preços administrados e os alimentos a domicílio do índice completo [54]. No relatório de dezembro de 2009 [5] o Banco central introduziu uma nova medida, segundo Silva Filho e Figueiredo [54], baseada em uma versão inicial de Silva Filho e Figueiredo [53]. Este último apresenta algumas críticas ao uso dos núcleos de inflação até então calculados. Uma delas é que o cálculo de EX0 remove itens que não deveriam ser removidos e que deve ser feito um estudo preliminar de quais itens remover, baseado nas suas volatilidades. O resultado é a proposta de um novo núcleo por exclusão (EX1) a ser calculado, removendo-se 12 itens, listados em [54] e [53]. O EX1 é o IPCA expurgado dos itens “consistentemente mais voláteis entre janeiro/1995 a julho/2007” [54, p. 9]. Mais tarde o EX1 foi atualizado para o IPCA expurgado dos itens mais voláteis entre agosto/1999 e junho/2013, obtendo-se o EX2 [54] que será abordado neste trabalho. Os itens que são excluídos do IPCA EX2 podem ser encontrados no Apêndice.

**Núcleo por médias aparadas (MA).** Bryan e Checchetti [10] propõem uma alternativa ao núcleo de inflação por exclusão. Supondo que a cada período observado, as maiores e menores variações nos preços dos itens que constituem o índice de inflação são ruídos, a proposta é remover sempre estes componentes. Diferente dos núcleos por exclusão, o núcleo por médias aparadas não exclui sempre os mesmos itens, mas exclui os que possuem variações relativamente maiores e menores em cada período. O ponto de referência desta exclusão é dado pela mediana (em vez da média) da amostra de variações, i.e. são

---

<sup>2</sup>*Consumer Price Index*

excluídos  $\alpha\%$  dos itens com variação menor do que a mediana e  $\alpha\%$  dos itens com variação maior. Bryan e Checchetti [10] justificam o uso da mediana por causa da tendência destas variações assumirem distribuições leptocúrticas (caudas pesadas) e a maior eficiência da mediana nestes casos. O Banco Central do Brasil divulga duas medidas de núcleo por médias aparadas, com e sem suavização. A que será utilizada é a medida por médias aparadas sem suavização (MA) a qual exclui no cômputo mensal de cálculo os itens com as 20% maiores e menores variações [54, 5].

**Núcleo por dupla ponderação (DP).** A última medida de núcleo oficial a ser avaliada neste trabalho é a medida por dupla ponderação calculada pelo Banco Central. Sua metodologia é apresentada no Relatório de Inflação de dezembro de 2009 [5]. Ela é conceitualmente parecida com a medida por médias aparadas, ao invés vez de excluir itens, ela recalcula os pesos, diminuindo os pesos dos itens com volatilidade mais alta. Esta volatilidade é dada pelo desvio padrão das variações de cada item em uma janela de 48 meses. Os pesos são então reponderados proporcionalmente ao inverso desta volatilidade.

Neste trabalho será estudada uma forma alternativa de medição do núcleo de inflação. Dowd, Cotter e Loh [18] sugerem que a medida de núcleo de inflação baseadas em wavelets são capazes de executar as duas tarefas esperadas de um núcleo: isolar a tendência da inflação de ruídos transitórios e possuir capacidade preditiva da inflação futura. Os autores constroem uma medida de núcleo baseada na *Discrete Wavelet Transform* (DWT) e obtêm sucesso, quando comparado a outras medidas, em obter uma estimativa de um núcleo confiável para os Estados Unidos.

Conforme apresentado no capítulo anterior, a DWT decompõe um sinal qualquer em resoluções temporais, da mais refinada a mais bruta. Pode-se ver isso como conjuntos de informação: as resoluções mais baixas<sup>3</sup> mostram mais informação acerca de um sinal, enquanto que resoluções mais altas representam menos informação. O trade-off é que resoluções baixas mostram apenas o que acontece com um sinal em períodos curtos de tempo, enquanto que resoluções mais altas, mostram o que acontece com o sinal em períodos longos de tempo. A análise multi resolução (MRA) assim obtida é um conjunto

---

<sup>3</sup>Em um computador, por exemplo, resolução baixa significa menos informação por área mostrada em um monitor. Com wavelets, resolução baixa significa mais informação localizada.

de informação sobre o sinal e pode-se escolher quais resoluções serão utilizadas para reconstruir um sinal estimado. Em geral escolhe-se remover as resoluções mais baixas, isto é, aquelas que representam acontecimentos de curta duração que aparecem diversas vezes em um sinal observado. Dowd, Cotter e Loh [18] utilizam essa premissa para reconstruir uma estimativa de núcleo de inflação sem os ruídos de curto prazo, i.e. uma estimativa composta por movimentos de médio e longo prazo.

Os problema de se utilizar a DWT em séries estatísticas é a incapacidade dela lidar com rotações de um sinal, isto é, ela é influenciada pela data inicial das observações. Além disso ela requer que a série temporal seja divisível por dois tantas vezes quanto forem realizadas decomposições em resoluções temporais diferentes. Para contornar esse problema, Anderson[1], Baqaee [2]; Denardin, Kozakevicius e Schmidt [16]; Plessis, Rand e Kotzé [47] e Lahura e Vega [40] utilizam a *Maximum Overlap Discrete Wavelet Transform* (MODWT) de Percival e Walden [46], que é capaz de realizar uma MRA similar a DWT mas não é influenciada pela localização das observações e funciona para qualquer número de observações.

O produto final de uma aplicação da MRA baseada na MODWT são decomposições de um sinal em escalas chamadas de *smooth* (**S**) e *detail* (**D**) que representam a parte de longo e curto prazo de um sinal, dada uma resolução escolhida. O primeiro nível da MRA,  $\mathbf{S}^1$  e  $\mathbf{D}^1$ , divide, aproximadamente, o espectro do sinal em componentes com frequência entre 0 e  $1/4$ , e  $1/4$  e  $1/2$  da taxa de observação, respectivamente, onde a taxa de observação é a periodicidade da amostra. Para o caso da inflação, essa taxa geralmente é mensal, portanto a parte *smooth* do primeiro nível da MRA representa acontecimentos que ocorrem com frequência de zero vezes ao mês, que é identificada como a média do sinal, até acontecimentos com periodicidade de quatro em quatro meses, ou um quarto de um ciclo em um mês. A parte *detail* representa acontecimentos com ciclos entre 2 e 4 meses. Nota-se que pelo Teorema de Shannon-Nyquist, ciclos de dois meses são a máxima frequência possível de se extrair de um sinal observado mensalmente.

O segundo nível da MRA divide o espectro da parte *smooth* do nível anterior pela metade, resultando em novos níveis *smooth* e *detail*, agora representando frequências entre 0 e  $1/8$ , e  $1/8$  e  $1/4$  ciclos por mês. Os níveis subsequentes são divisões do espectro do nível *smooth* anterior pela metade. Isso resulta em aproximações mais detalhadas do

sinal pois são ganhas mais informações acerca do espectro. A equação do MRA para um sinal  $x$  qualquer é dada por:

$$x = \mathbf{S}^1 + \mathbf{D}^1 = \mathbf{S}^2 + \mathbf{D}^2 + \mathbf{D}^1 = \dots = \mathbf{S}^J + \mathbf{D}^J + \mathbf{D}^{J-1} + \mathbf{D}^{J-2} + \dots + \mathbf{D}^1. \quad (4.1)$$

Para aplicações ao estudo do núcleo de inflação, uma MRA feita até certo nível  $J$  isola componentes com de frequência mais baixa, no intervalo  $[0, 1/2^{J+1}]$  dos demais, incluindo componentes que tem ciclo longo e duradouro, ou seja, a tendência de longo prazo da inflação. A parte devida às frequências mais altas é identificada como ruído<sup>4</sup>. Está é a heurística do trabalho de Dowd, Cotter e Loh [18], que procura construir uma MRA de um nível grande o suficiente para isolar o que é tendência do que são flutuações. Os trabalhos posteriores (referências [16],[2], [1],[40] e [47]) que aplicaram as ferramentas wavelet para o problema seguiram a mesma linha.

A descrição da metodologia dada nos parágrafos acima está de acordo com a exposição dada na seção 3.2. Como foi mencionado, embora seja possível dizer que os valores da MODWT formam um vetor com componentes de frequência pertencentes aos intervalos mencionados, não se pode afirmar que estes componentes se apresentam uniformemente em todas as partes do sinal, nem que eles estão presentes em um instante qualquer. Além disso, se desejarmos interpretar a MODWT como um filtro, será necessária a hipótese de periodicidade - uma suposição certamente irrealista no contexto de séries temporais de variáveis econômicas.

Por outro lado a MODWT tem resolução temporal conhecida, então pode-se afirmar que cada um de seus valores tem origem certa no tempo. Ao invés de falar em frequências, os parágrafos anteriores poderiam falar em médias e diferenças da série temporal e o equivalente a filtrar seria isolar os valores que formam a tendência da série em cada instante dos valores que representam oscilações bruscas. Resoluções baixas signifi-

---

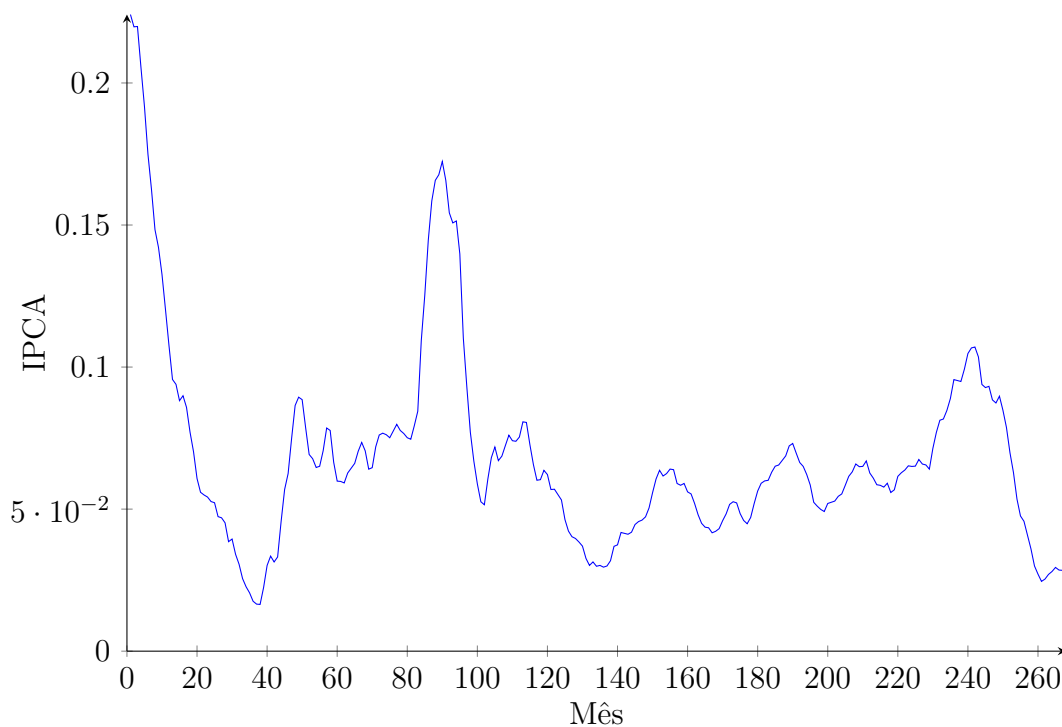
<sup>4</sup>Este nome é um legado da terminologia de processamento de sinais. As frequências mais altas estão ligadas a acontecimentos transitórios que afetam o nível de preços, como choques de oferta ou demanda de baixa duração. Eles não são necessariamente interferências nas observações, mas parte do processo econômico. Porém como o objetivo de se calcular o núcleo de inflação é possibilitar medidas de controle inflacionário de médio e longo prazo, desconsiderando flutuações momentâneas, a mesma aplicação que serve para remover artefatos indesejados de um sinal físico, supostamente serve para remover componentes transitórios da inflação.

cam médias e diferenças tomadas em intervalos curtos, enquanto que resoluções altas são tomadas em intervalos longos. Pode-se dizer que a remoção de níveis resulta em remoção de componentes oscilatórios de curto prazo.

## 4.2 Os Dados

Para reproduzir os exercícios de obtenção das estimativas de núcleo de inflação baseadas em wavelets, sigla WIM, serão utilizados os dados primários obtidos através do Sistema de Recuperação Automática (SIDRA) do Instituto Brasileiro de Geografia e Estatística (IBGE) [36] e também as estimativas de núcleo oficiais calculadas pelo BACEN, disponíveis no Sistema de Séries Temporais da instituição [4]. Os dados serão utilizados na forma como são obtidos, isto é, sem ajustes sazonais, mas serão apresentadas, a não ser que dito o contrário, como taxas anualizadas. Enquanto Dowd, Cotter e Loh [18] utilizam dados dessazonalizados, os outros trabalhos mencionados obtêm WIMs baseadas nos dados brutos.

**Figura 4.4:** IPCA mensal a taxas anualizadas entre dez/1995 e mar/2018



Fonte: IBGE [36]

A Figura 4.4 mostra o gráfico da inflação brasileira medida pelo IPCA, de dezembro de 1995 até março de 2018, mostradas em valores decimais (i.e.  $20\% = 0.2$ ). Embora os dados existam desde a década de 1960, a alta magnitude dos valores nos períodos anteriores, especialmente no períodos de hiperinflação da década de 1980, tornam a análise mais difícil, especialmente por que a variância da série aumenta muito. A partir do plano Real (1994) a inflação se estabiliza. Além disso, como as medidas de núcleo calculadas pelo banco central não vão até o começo da amostra do IPCA, foi escolhido começar a série pelo fim de 1995, isto é, calcular a primeira taxa anualizada possível com os dados.

A metodologia de cálculo atual do IPCA é descrita em IBGE [35]. Este índice abrange famílias que ganham de um a 40 salários mínimos e é calculado mensalmente. O motivo da banda de salários pesquisada é o fato de que o consumo de famílias fora delas, segundo o IBGE [35], é instável e difícil de medir. Porém, existe o Índice Nacional de Preços ao Consumidor (INPC), que calcula o índice de preços baseado no consumo das classes mais baixas de renda (de um a cinco salários mínimos). De qualquer forma, o índice oficial, ao menos para fins de meta de inflação divulgadas, é o IPCA. O objetivo do índice é verificar a mudança no custo de vida nominal dos consumidores em determinadas áreas geográficas (ver a lista em IBGE [37]). Ele consiste em uma média ponderada dos gastos da população objetivo, verificados na pesquisa de orçamento familiar [35], isto é, o gasto relativo ao gasto total das famílias em cada bem ou serviço. A cada mês é feita uma pesquisa dos preços pertinentes em estabelecimentos cadastrados, calculada a variação e realizado o agrupamento por classe (subitens). A partir dos subitens é feito o cálculo dos gastos nos itens, por média ponderada, com pesos dados pelos gastos relativos em cada produto. A partir dos itens é calculado o índice. Este por sua vez é um número que representa o dispêndio relativo com a cesta de consumo entre diversas datas. Isto é, escolhendo uma data  $t_0$ , onde o índice é 100, verifica-se qual o custo de consumir a cesta padrão<sup>5</sup> na data  $t_j$  (que pode ser antes ou depois de  $t_0$ ), por exemplo, se o custo aumentou 20%, o novo índice é 120. Como para políticas públicas o interesse é a taxa de inflação, os dados utilizados são as variações percentuais mensais deste número índice.

---

<sup>5</sup>Nota-se que a cesta padrão muda de período para período.



### 4.2.1 O modelo

Supomos que a inflação é dada por um componente chamado núcleo e um componente de ruído aleatório,

$$\pi_t = \pi_t^c + \epsilon_t \quad (4.2)$$

onde  $\pi_t$  é a inflação observada na data  $t$ ,  $\pi_t^c$  é o núcleo de inflação que estimamos, mas que deve ser exógeno ao modelo, e  $\epsilon_t$  é uma variável aleatória que se supõe independente e identicamente distribuída, com distribuição normal e média zero. A extração de sinal procura obter  $\pi_t^c$  a partir dos dados brutos.

## 4.3 Medidas de Núcleo baseadas em Wavelets

Para construir as medidas WIM no caso brasileiro Denardin, Kozakevicius e Schmidt [16] executam diversas MRAs, baseadas na MODWT de diversas wavelets diferentes. Os dados utilizados são o IPCA e para estimar o núcleo os autores excluem os níveis *detail* de um a três, isto é, excluem os componentes do sinal com ciclos entre 2 e 16 meses, supondo que o núcleo de inflação esteja presente nos ciclos com duração mais baixa do que 16 meses, em conformidade com Lahura e Vega [40], que consideram a remoção desses níveis suficiente para remover a maior parte dos componentes transitórios indesejados.

Existe uma outra razão prática para não excluir muitos níveis: as wavelets fazem uso da hipótese de periodicidade de um sinal e à medida que se sobem os níveis é necessária a extensão dos dados para além do número de observações, proporcionalmente ao suporte da wavelet utilizada. Isso pode ser um problema, pois wavelets com suporte grande vão precisar de muito mais dados e por consequência mais extensão do sinal. Para se ter uma ideia, são necessários 76 valores de um sinal para calcular os coeficientes da MODWT no nível quatro da DB6. Segundo Percival e Walden [46] isso significa 76 coeficientes que devem ser descartados para usos estatísticos. Considerando que o IPCA tem 268 observações, isto é um problema. Portanto a escolha das wavelets utilizadas, bem como o nível máximo de decomposição, são importantes. Embora no trabalho de Denardin, Kozakevicius e Schmidt [16] sejam analisadas várias wavelets diferentes, da família Daubechies e da família Symlets, apontando que todas elas são capazes de criar

uma medida adequada, as Daubechies 1 e 2 (isto é a wavelet de Haar e a DB2) já possuem todas as qualidades necessárias. Além disso elas são as duas menores Wavelets possíveis, e portanto são as mais parcimoniosas e requerem menor uso da hipótese de periodicidade. em um estudo preliminar, estimamos o núcleo de inflação utilizando a wavelet Haar e a wavelet DB2. Pelos critérios adotados, a DB2 apresentou melhores resultados. Portanto, no que segue, apresentamos somente os resultados relacionados à DB2.

As medidas baseadas em wavelets que serão calculadas serão (i) com exclusão dos três primeiros níveis *detail*, (ii) com exclusão dos quatro primeiros níveis *detail*, e também (iii) medidas com uma etapa intermediária de *thresholding*.

Quando o núcleo é calculado removendo-se níveis de detalhe, é feita a suposição de que os componentes de um sinal atribuídos a ruídos estão pelo menos majoritariamente contidos naqueles níveis de detalhe removidos. Podem haver oscilações tão grandes, mesmo que transitórias, que mudam a tendência de um sinal, acontecimentos que são chamados de quebra de tendência. Quando elas acontecem, é assumido que parte do sinal estará presente nos detalhes dos níveis mais baixos e portanto eles não devem ser removidos. Donoho e Johnstone [17] utilizam a metodologia de *thresholding* nas transformadas wavelets para isolar, nos níveis mais baixos o que é sinal e o que é ruído. Um dos resultados de seu trabalho é provar que o *soft thresholding* é capaz de fazer isso melhor do que simplesmente remover níveis, sob a hipótese de que o sinal subjacente segue um modelo determinístico, uma suposição talvez inadequada para modelos econômicos. O seu produto é uma equação similar a (4.1):

$$x = \mathbf{S}^1 + \mathbf{D}^{*1} = \mathbf{S}^2 + \mathbf{D}^{*2} + \mathbf{D}^{*1} = \dots = \mathbf{S}^J + \mathbf{D}^{*J} + \mathbf{D}^{*J-1} + \mathbf{D}^{*J-2} + \dots + \mathbf{D}^{*1}. \quad (4.3)$$

onde o sobrescrito \* denota que o nível sofreu um processo de *thresholding*. Nota-se que apenas os níveis *detail* sofrem *thresholding*. O *soft thresholding* consiste em eliminar os coeficientes abaixo de um certo nível (o *threshold*) e reduzir, em um valor igual ao *threshold*, os coeficientes que se encontram acima deste nível. Em geral esse processo gera vetores esparsos, i.e. com poucas entradas [38]. Este procedimento é menos arbitrário que remover os níveis *detail* mais baixos, mas envolve certa arbitrariedade quanto ao nível  $J$  até o qual será calculada a transformada wavelet. A extensão do algoritmo de *thresholding* para a MODWT é dada por Percival e Walden [46]. O cálculo é o mesmo utilizado para calcular as transformadas wavelets anteriores, mas com a etapa intermediária de

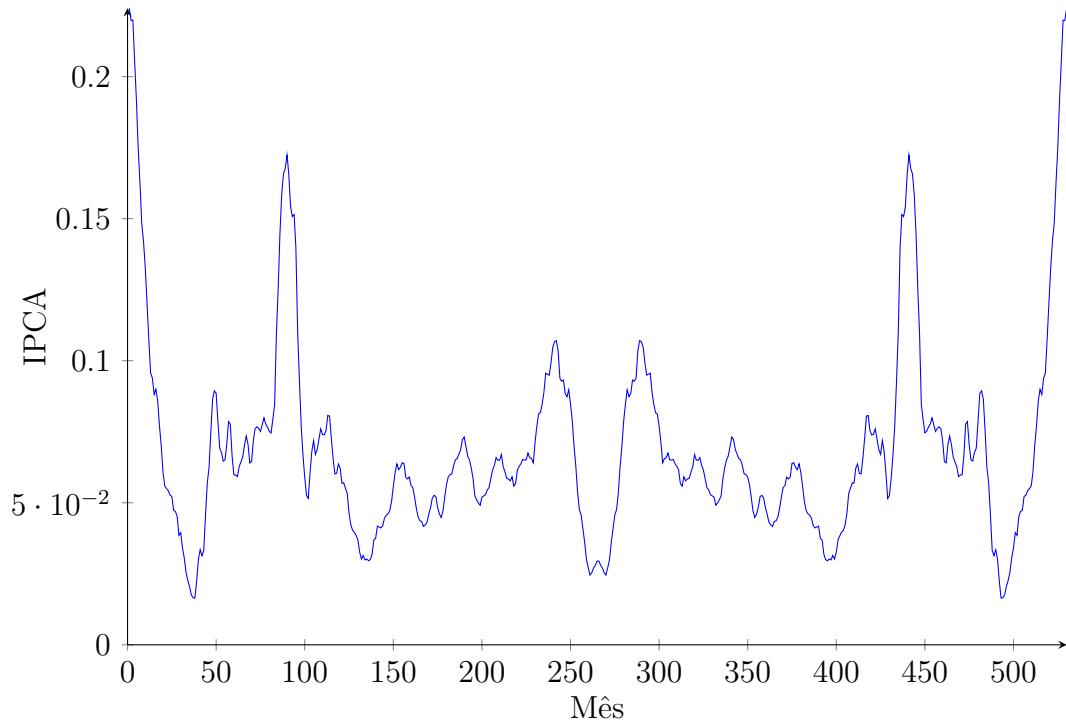
*thresholding* antes da inversão do processo. Para ser aplicado, o método supõe que um sinal segue o modelo em (4.2). Onde  $\pi_t^c$  representa o sinal a ser extraído e  $\epsilon_t$  o ruído que deve ter distribuição normal com média 0 e variância constante. O uso da MODWT com *thresholding* para isolar o núcleo de inflação foi utilizado primeiramente por Baqaee [2] para a inflação neozelandesa. Neste trabalho serão calculados MRA's com *thresholding* até o terceiro e quarto níveis.

Inicialmente, quatro WIM's serão calculadas, duas com remoção de níveis e duas com *thresholding*. Para facilitar a notação, será utilizada a seguinte nomenclatura: o nome da wavelet, DB2, seguido do nível até o qual os componentes *detail* foram removidos ou sofreram *thresholding*, DB23 ou DB24, seguido da letra t para denotar se foi realizado o *thresholding*. Assim DB24t é a medida de núcleo baseada em uma MRA com a Daubechies 2 com *thresholding* até o quarto nível. Para o período analisado, junto das medidas baseadas em wavelets também serão analisadas duas medidas reportadas pelo BACEN, a EX2 e a DP, pois a MA só possui dados a partir de 1999.

O primeiro passo para construção do núcleo baseado em wavelets é definir como será estendido o sinal. Geralmente é feita a hipótese de periodicidade, isto é, se o sinal tem tamanho  $N$  então ele é estendido de forma que  $x_{t+k} = x_{(t+k) \bmod N}$  se  $k \geq N$ . A desvantagem de se estender um sinal dessa forma é que ele é sujeito a descontinuidades importantes. Para o IPCA estendido periodicamente, se  $t = 268$  (a última observação disponível) o valor observado será de 0.027 enquanto que em  $t = 269$  ele será 0.224, um salto de mais de 800%. Uma solução possível é apontada por Broughton e Bryan [9] e consiste em estender o sinal de forma que se  $t > N$  então  $x_t = x_{N-(t-N)}$ . Essa forma de extensão, que chamaremos de extensão por reflexão, elimina descontinuidades no sinal e é possivelmente mais condizente com a realidade. A Figura 4.5 mostra o IPCA estendido por reflexão.

Para obter a MRA não é necessário estender todo sinal como na figura 4.5, pois apenas algumas entradas vão precisar dos dados estendidos. Para cada nível, Percival e Walden [46] fornecem uma fórmula para saber quantas entradas estendidas são necessárias. De modo a ganhar um pouco de eficiência computacional, só são estendidas as entradas necessárias, dependendo do nível máximo de cada MRA. Os algoritmos de cálculo das MRAs foram escritos pelo autor deste trabalho, para o software Scilab, e fortemente baseados nas referências [9], [58] e [46]. O algoritmo original é o piramidal de Mallat

**Figura 4.5:** IPCA Estetendido por Reflexão



Fonte:elaborado pelo autor, baseado nos dados do IBGE [36]

[43]. Os valores dos filtros wavelet foram obtidos no website Wavelet Browser [59] e redimensionados para a MODWT de acordo com Percival e Walden [46].

A figura 4.6 mostra os gráficos das medidas de núcleo e do IPCA. Nota-se que enquanto o IPCA tem várias oscilações, as medidas baseadas em wavelets são mais suaves, e o IPCA parece oscilar em torno delas. Isso é uma característica da estimação de sinais com wavelets: elas suavizam o gráfico de um sinal. A tabela 4.1 apresenta algumas estatísticas descritivas do IPCA e das medidas de núcleo. Também é mostrado na última linha o p-valor do teste de Jarque-Bera para normalidade da diferença entre IPCA e núcleo, isto é, assumindo que estamos estimando um sinal, removendo um termo de ruído com distribuição normal.

O teste de normalidade utilizado por Baqaee [2] tem o intuito de verificar quais das wavelets utilizadas saem-se bem removendo ruídos, pois a hipótese do *thresholding* é de que os termos de ruído tem distribuição normal. Deve ser salientada a diferença entre as motivações do teste de normalidade de Baqaee e do teste realizado por Cotter e Dowd [18]. Neste os autores supõe que a remoção de níveis *detail* só deve ser feita

até o ponto que estes tenham distribuição normal, isto é, eles não removem componentes que tenham distribuição normal. Naquele a motivação é oposta, pois deseja-se remover componentes com distribuição normal pois estes são supostamente os ruídos que se deseja eliminar. Ainda notamos que os núcleos calculados pelo Banco Central não tem por pressuposto a separação de um sinal dos ruídos, então não é esperado a priori que a diferença entre IPCA e estas medidas de núcleo tenha distribuição normal.

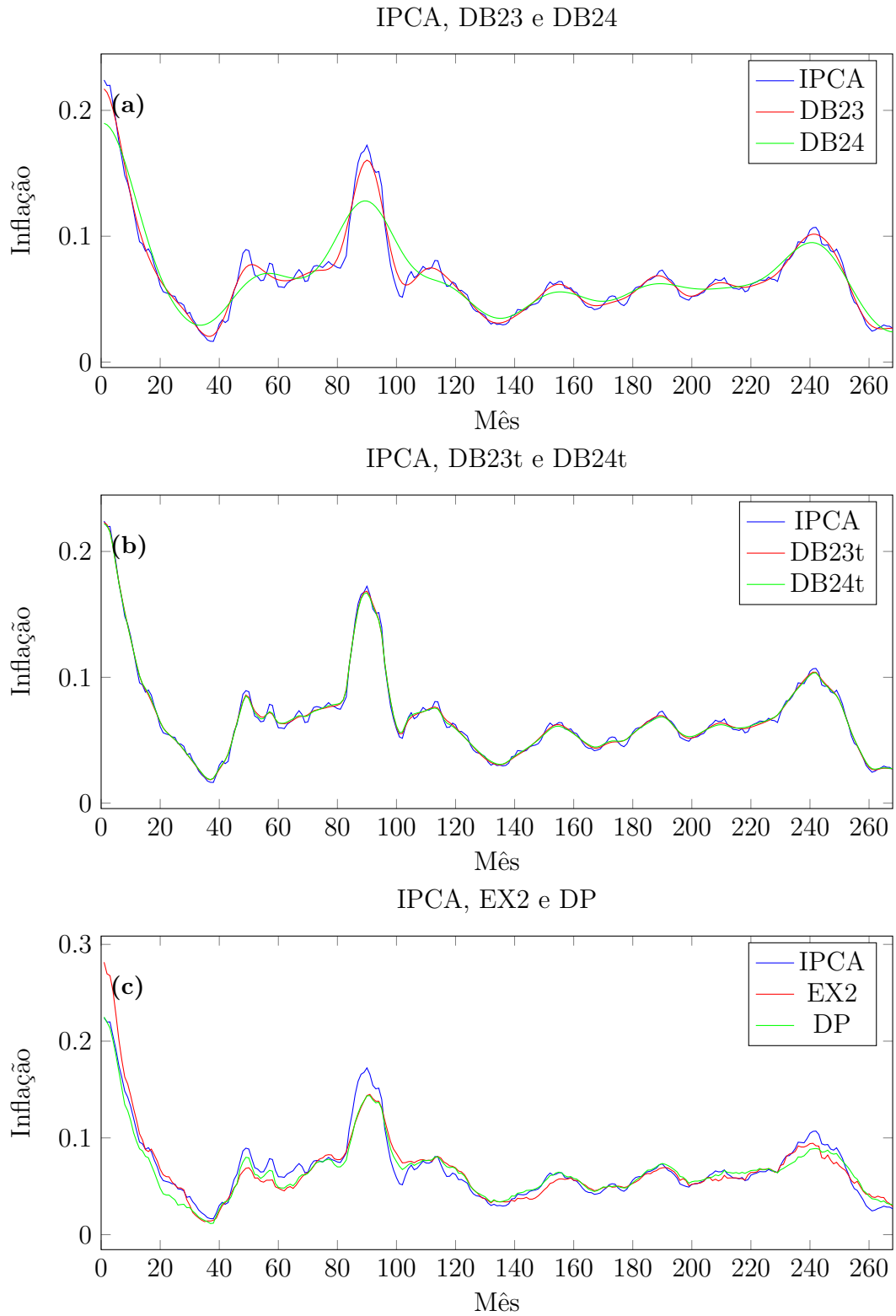
**Tabela 4.1:** Estatísticas descritivas

	<b>IPCA</b>	<b>DB23</b>	<b>DB24</b>	<b>DB23t</b>	<b>DB24t</b>	<b>EX2</b>	<b>DP</b>
<b>Média</b>	0.0687	0.0687	0.0687	0.0687	0.0687	0.0682	0.0664
<b>D.P.</b>	0.0356	0.0344	0.0311	0.0352	0.0348	0.0385	0.0319
<b>Variação Total</b>	1.0438	0.7400	0.5153	0.8240	0.8048	0.8647	0.8292
<b>p-valor J.B.</b>	-	0.0000	0.0000	0.7279	0.1629	0.0000	0.0000

Fonte: Elaborado pelo autor. Baseado em dados do Banco Central do Brasil [3]

Algumas informações da tabela 4.1 são importantes. Primeiro, a média amostral do IPCA e do núcleo deve ser parecida, o que acontece com todas as medidas. Isso, para o caso das WIM's deve-se ao próprio método, pois a decomposição em wavelets mantém a média do sinal. Uma outra qualidade esperada de um núcleo é a redução da volatilidade, medida pelo desvio padrão. Neste caso, apenas a medida EX2 possui desvio padrão maior que o IPCA. Todos os núcleos estimados possuem variação total menor do que o IPCA. Quanto ao teste de normalidade, as medidas que passam com significância de 1% são aquela nas quais foi adotada a etapa de *thresholding*.

**Figura 4.6:** IPCA e Núcleos de Inflação



Fonte: Banco Central do Brasil [4].

De forma a avaliar a adequação das medidas de núcleo acima introduzidas são realizados 3 testes estatísticos, sumarizados por Rich e Steindel [24], que também impõem critérios qualitativos não abordados aqui, e Santos e Castelar [50]. Em geral os critérios satisfazem pelo menos uma das visões do que é núcleo de inflação. A primeira visão, de Checetti [13], que considera o núcleo como a tendência subjacente da inflação. Isso significa que ela deve ser uma medida não enviesada da inflação e considerando que o ruídos devam ter média zero, deve satisfazer:

$$E(\pi_t | \pi_t^c) = \pi_t^c \quad (4.4)$$

Para que isso aconteça segundo Santos e Castelar [50], desde que ambas as séries sejam estacionárias, deve ser feita a seguinte regressão:

$$\pi_t = \alpha + \beta \pi_t^c + \epsilon_t \quad (4.5)$$

e testado com um teste F para as restrições conjuntas  $\alpha = 0$  e  $\beta = 1$ . A parte das regressões deste trabalho é realizada com os softwares Eviews<sup>6</sup> e RStudio<sup>7</sup>. O motivo da utilização dos dois programas é a facilidade do uso do Eviews, e a capacidade de executar análises com um número grande de séries diferentes ao mesmo tempo do R. O *package* em R de todas as funções utilizadas, bem como os parâmetros passados serão mencionados no rodapé. Para nomear as funções o texto será usada a fonte `typewriter`.

Os testes de estacionariedade foram realizados sem constante nem tendência, com a função `adfTest`<sup>8</sup> [62] do R, que é o teste aumentado de Dickey-Fuller. A hipótese nula de não estacionariedade é rejeitada para todas as séries, com p-valor menor do que 1%, portanto não serão reportados os p-valores individualmente.

Os resultados da estimação da regressão (4.5) por mínimos quadrados (OLS) são apresentados na tabela 4.2, aproximados por quatro casas decimais. As regressões foram feitas com a função `lm`<sup>9</sup> do R. O teste F para restrição foi calculado pelo autor seguindo Gujarati e Porter [29, p. 263]. Sucintamente, este teste consiste em verificar se a soma

---

<sup>6</sup><http://www.eviews.com/home.html>

<sup>7</sup><https://www.rstudio.com/>

<sup>8</sup>Package: `fUnitRoots`, parâmetros: `x = série` e `type = "nc"`

<sup>9</sup>Package: `stats`, parâmetros: `IPCA ~ Núcleo`

dos quadrados dos resíduos das regressões com e sem restrições diferem substancialmente. A estatística deste teste, assume, sob hipótese nula de que a restrição é verdadeira, distribuição F com (2, 266) graus de liberdade. O valor apresentado na tabela é o p-valor do teste.

**Tabela 4.2:** Teste de viés

	<b>DB23</b>	<b>DB24</b>	<b>DB23t</b>	<b>DB24t</b>	<b>EX2</b>	<b>DP</b>
$\alpha$	-0.0017	-0.0056	-0.0007	-0.0013	0.0087	-0.0032
<b>D.P.</b> $\alpha$	0.0007	0.0017	0.0003	0.0003	0.0013	0.0012
<b>p-valor</b>	0.0098	0.0015	0.0431	0.0002	0.0000	0.0062
$\beta$	1.0251	1.0811	1.0097	1.0186	0.8804	1.084
<b>D.P.</b> $\beta$	0.0086	0.0231	0.0043	0.0044	0.017	0.0159
<b>p-valor</b>	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
<b>teste F</b>	0.0156	0.0024	0.0784	0.0002	0.0000	0.0000

Fonte: elaborado pelo autor, baseado em dados do Branco Central [4].

Como mostra a tabela 4.2, o beta das medidas varia de 0.8804 da EX2 para 1.0840 da DP. Os betas das WIM's parecem estar próximos de 1. Todos interceptos são da ordem de  $10^{-3}$ . Claro que sozinhos os coeficientes não falam muito, pois o teste requer que a restrição  $\beta = 1$  e  $\alpha = 0$  seja estatisticamente plausível, e apenas duas medidas seriam aceitas, a 1% (DB23 e DB23t) e 1 medida aceita a 5% (DB23t). Porém este teste pode ser melhorado. O teste F aplicado só leva em conta os resíduos estimados das regressões, mas existe evidência delas terem problemas de correlação serial, fato constatado apos um teste de Breusch-Godfrey para autocorrelação realizado no Eviews. Neste caso, os coeficientes são não viesados [61], mas a variância dos coeficientes, e o teste F são inválidos. Para corrigir este problema, as regressões foram refeitas no Eviews com o estimador de Newey-West (ver [19]) para a covariância. Este, também chamado de HAC, é robusto na presença de auto correlação e heterocedasticidade desconhecidos [61]. É importante mencionar que ele não modifica os coeficientes obtidos da estimação por mínimos quadrados, por isso, para refazer o teste de restrição, o teste F não é adequado, já que os erros estimados



continuam iguais. Em vez disso é feito o teste de Wald para restrição nos coeficientes. Este leva em conta a nova matriz de covariância dos coeficientes [22]. A tabela 4.3 mostra os p-valores do teste de Wald, para a hipótese nula de que a restrição é verdadeira.

**Tabela 4.3:** Teste de Wald para restrição nos coeficientes

	<b>DB23</b>	<b>DB24</b>	<b>DB23t</b>	<b>DB24t</b>	<b>EX2</b>	<b>DP</b>
Teste de Wald	0.1764	0.2462	0.0362	0.0002	0.0555	0.0300

Fonte: elaborado pelo autor, baseado em dados do Branco Central [4]

Neste caso todas as medidas exceto a DB24t passam no teste a 1%. Isso é um indício de que as medidas de núcleo conseguem remover os ruídos da inflação e continuar a medir a inflação média, ou seja, são boas medidas de núcleo na visão de Chechetti.

A segunda visão do que é um núcleo adequado é dada por Blinder [7], que requer que o núcleo de inflação seja capaz de prever a inflação futura melhor do que outras estatísticas. Entre os testes possíveis para verificar isso está o teste de ajuste dinâmico, que pode ser feito de diversas formas (ver referências[24], [50], [2]). Um teste relativamente simples é proposto por Cogley [15, p. 104] <sup>10</sup> que consiste em verificar o ajustamento através do tempo entre inflação e núcleo. Por hipótese, a inflação a qualquer instante pode ser maior ou menor que seu núcleo, mas após tempo suficiente passar, a expectativa é que ela se reaproxime. O teste de Cogley realiza a seguinte regressão:

$$\pi_{t+h} - \pi_t = \alpha + \beta(\pi_t - \pi_t^c) + \epsilon_t \quad (4.6)$$

Onde por suposição  $\alpha$  deve ser igual a zero, já que uma boa medida de núcleo satisfaz (4.4). Segundo Cogley [15], para  $h$  suficientemente grande  $\pi_{t+h} - \pi_t$  e  $\beta(\pi_t - \pi_t^c)$  devem ser inversamente proporcionais com  $\beta$  próximo de -1. Este teste pode ser visto como um indicador de para onde a inflação vai em  $h$  meses. Se ela estiver acima (abaixo do núcleo) em um dado  $t_0$ , esperamos que a inflação futura caia (suba) em relação a inflação hoje. Denardin, Kozakevicius e Schmidt [16] usam um teste (4.6) diferente, proposto por

<sup>10</sup>Deve se mencionar que Cogley [15] trabalho propõe uma outra medida de núcleo de inflação baseada em métodos no domínio da frequência.

Mehra e Reilly (2009, apud [16])<sup>11</sup> que leva em conta também *lags* da inflação. Uma razão para usar lags é que em testes preliminares estimação de (4.6) parece ter problemas de especificação, algo observado por Silva Filho e Figueiredo [53], e a inclusão dos *lags* ajuda a obter uma estimativa melhor. Mesmo assim, nas regressões feitas desta forma, os coeficientes ficam muito afastados do ideal, e ainda há indícios de má especificação do modelo, com erros autocorrelacionados e mesmo com métodos robustos para a covariância, as regressões são inconclusivas, com intervalos de confiança grandes. Para corrigir isso, seguimos o tratamento de Wooldridge [61, p467 - 73] e do manual do Eviews [20] para erros auto correlacionados e utilizamos um estimador de mínimos quadrados generalizados (GLS). Diferente do método de Newey-West, estes modelos modificam os coeficientes das regressões e devem ser usados com cautela. Segundo Wooldridge [61], em geral, se os coeficientes estimados por OLS e GLS forem muito diferentes, deve se preferir o OLS. Porém se valores dos coeficientes obtidos pelo GLS forem mais condizentes com a teoria, estes podem ser utilizados.

Para realizar as regressões foi utilizada a função `auto.arima`<sup>12</sup> [64, 33] do R para  $h = 6, 12$  e  $18$ . Esta função escolhe automaticamente um modelo para a equação (4.6), entre modelos com erros seguindo processos ARIMA. O critério de seleção (AIC, AICC ou BIC) pode ser escolhido pelo usuário, bem como parâmetros máximos e mínimos para o número de lags dos processos ARIMA. Neste trabalho foi utilizado os parâmetros *default*, que podem ser vistos na documentação do R, em Hyndman et al. [33]. A tabela 4.4 mostra os resultados do teste de ajuste dinâmico. Notamos que os modelos automaticamente escolhidos em geral não possuem intercepto, portanto os interceptos não são reportados.

**Tabela 4.4:** Teste de ajuste dinâmico

	<b>DB23</b>	<b>DB24</b>	<b>DB23t</b>	<b>DB24t</b>	<b>EX2</b>	<b>DP</b>
	<b>h = 6</b>					
$\beta$	-1.0262	-1.0426	-1.1543	-1.1877	-0.7931	-1.1652
<b>D.P.</b> $\beta$	0.0669	0.0643	0.0822	0.0865	0.1073	0.1307
<b>ARMA</b>	(2, 2)	(2, 3)	(2, 0)	(2, 0)	(2, 2)	(2, 0)

<sup>11</sup>MEHRA, Y. P.; REILLY, D. Short-term headline-core inflation dynamics. *Economic Quarterly*, Federal Reserve Bank of Richmond, v. 95, n. 3, p. pp. 289–313, Summer 2009.

<sup>12</sup>Package forecast, parâmetros:  $y = \text{IPCA}$ ,  $xreg = \text{núcleo}$ .

	<b>h = 12</b>					
$\beta$	-1.4091	-1.4126	-1.5367	-1.5772	-1.2255	-1.8668
<b>D.P.</b> $\beta$	0.0596	0.0574	0.0803	0.0849	0.1208	0.1401
<b>ARMA</b>	(2, 0)	(1, 0)	(2, 1)	(2, 1)	(3, 2)	(2, 0)
	<b>h = 18</b>					
$\beta$	-1.0079	-0.9808	-1.1937	-1.3093	-0.7326	-1.3754
<b>D.P.</b> $\beta$	0.0654	0.0635	0.0804	0.0869	0.1098	0.1291
<b>ARMA</b>	(2, 2)	(2, 3)	(3, 2)	(5, 3)	(3, 3)	(2, 2)

Fonte: elaborado pelo autor, baseado em dados do Branco Central [4]

Neste caso não é necessário um teste de restrição para verificar se  $\beta = -1$  pois o objetivo deste teste é verificar o efeito de desvios entre inflação e núcleo na inflação futura. Para  $h = 6$  todas as medidas menos EX2 parecem se sair como previsto, com valores de  $\beta$  próximos de  $-1$ . Para  $h = 12$  o IPCA é mais responsivo, com todas as medidas, exceto EX2 (o desvio padrão é grande o suficiente para incluir  $-1$  no intervalo de confiança) confortavelmente abaixo de  $-1$ . Para 18 meses a repete-se a situação de 6 meses.

Os resultados deste teste mostram que o IPCA se ajusta ao núcleo através do tempo, especialmente os núcleos baseados em wavelets. Se a inflação estiver acima dele, ela tende a cair nos intervalos de tempo analisados, e se ela estiver abaixo, ela tende a subir. A utilidade deste teste para analisar a inflação é saber qual a tendência da inflação, tudo mais constante.

A última estatística de qualidade de uma medida de núcleo que será utilizada é a sua capacidade preditora de inflação futura. Para fazer isso Baqaee [2] estima um modelo preditor do núcleo, isto é, ele realiza uma regressão do tipo  $\pi_t^c = f(\pi_t^c)$  para tentar prever os valores futuros deste e compara com os valores efetivos da inflação, sendo que a melhor estimativa de núcleo por esse parâmetro é dada por aquela que tem a menor estatística raiz do erro quadrado médio, sigla *RMSE*, dada por:

$$RMSE_{t_0} = \sqrt{\frac{1}{12} \sum_{t=t_0}^{12} (\pi_t - \hat{\pi}_t^c)^2} \quad (4.7)$$

onde  $\pi_t$  é a inflação observada e  $\hat{\pi}_t^c$  é o núcleo previsto.

A equação (4.7) retorna a raiz do erro quadrado médio de previsão  $k$  passos a frente. Mas em vez de fazer apenas uma vez este teste para todos os dados faremos o procedimento de previsão da seguinte maneira, que é uma simplificação do teste realizado por Plessis, du Rand e Kotzé [47] para verificar quais medidas de núcleo previam melhor a inflação: dividimos a amostra em duas partes, a primeira com os primeiros  $2/3$  das observações e outra com o restante. Isso gera duas amostras distintas, uma com 179 observações e a outra com 89 observações. O primeiro passo é utilizar os dados das primeiras 179 observações para estimar modelos de regressão para os núcleos contra *lags* deles mesmos, i.e. estimar  $\pi_t^c = \alpha_1 \pi_{t-1}^c + \alpha_2 \pi_{t-2}^c + \dots$ . Em seguida é calculada a (4.7) para 12 passos a frente, com previsões feitas com cada modelo estimado, isto é, estimamos

$$\sqrt{\frac{1}{12} \sum_{t=180}^{191} (\pi_t - f(\alpha_1 \pi_{t-1}^c + \alpha_2 \pi_{t-2}^c + \dots))^2}.$$

Após estimados todos os modelos e RMSE's a amostra é acrescida de uma observação e são feitas novas regressões e computadas novas RMSE's. Depois a amostra é acrescida de outra observação, e repete-se o processo até chegar a amostra com 256 observações (as últimas 12 observações são utilizadas para calcular a última RMSE). Dessa forma são produzida 79 estatísticas de erro de predição da inflação para cada núcleo. Por fim toma-se a média delas para produzir uma estatística final da média das RMSE's.

O objetivo deste teste é verificar se alguém tentando prever a inflação o faria melhor prevendo *in lieu* o núcleo. Fazendo o teste diversas vezes com dados atualizados reproduz alguém atualizando seu modelo através do tempo, e podemos verificar qual das medidas de núcleo se sai melhor (ou pior), em média, prevendo a inflação. Para estimar cada modelo em cada etapa é novamente utilizada a função `auto.arima`<sup>13</sup> do R. Para fazer a previsão é utilizada a função `forecast`<sup>14</sup> e para produzir as estatísticas da raiz do erro quadrado médio é utilizada a função `accuracy`<sup>15</sup>. Para comparar a performance dos núcleos, em vez de compara-los entre si, eles são comparados com um *benchmark*, que é uma regressão da inflação contra *lags* dela mesma. O procedimento de obtenção das RMSE's permanece o mesmo, mas agora temos como previsão uma função da própria inflação ( $f(\pi_t, \pi_{t-1}, \dots)$ ). A tabela 4.5 reporta os resultados. Na primeira coluna

<sup>13</sup>Package `forecast`, parâmetros:  $y = \text{IPCA}$ ,  $x_{\text{reg}} = \text{núcleo}$ .

<sup>14</sup>Package `forecast`, parâmetros:  $\text{object} = \text{regressão}$ ,  $h = 12$ .

<sup>15</sup>Package `forecast`, parâmetros:  $f = \text{objeto } \text{forecast}$ ,  $x = \text{IPCA}$ .

é mostrada a média das RMSE's e na segunda a razão entre a média do núcleo e a média do *benchmark*.

**Tabela 4.5:** Teste de poder preditivo dos núcleos

	Média RMSE	Razão <i>Benchmark</i>
<b>IPCA</b>	0.0133	1.0000
<b>DB23</b>	0.0214	1.6106
<b>DB24</b>	0.0187	1.4078
<b>DB23t</b>	0.0149	1.1226
<b>DB24t</b>	0.0148	1.1192
<b>EX2</b>	0.0142	1.0709
<b>DP</b>	0.0147	1.1053

Fonte: elaborado pelo autor, baseado nos dados do Banco Central do Brasil [3]

Nenhuma das medidas de núcleo se saiu melhor do que o *benchmark* para prever a inflação, neste teste, com estes modelos. A que se saiu melhor foi a EX2 que previu a inflação com erro, em média, 7.09% maior que um modelo com a própria inflação foi capaz de fazer.

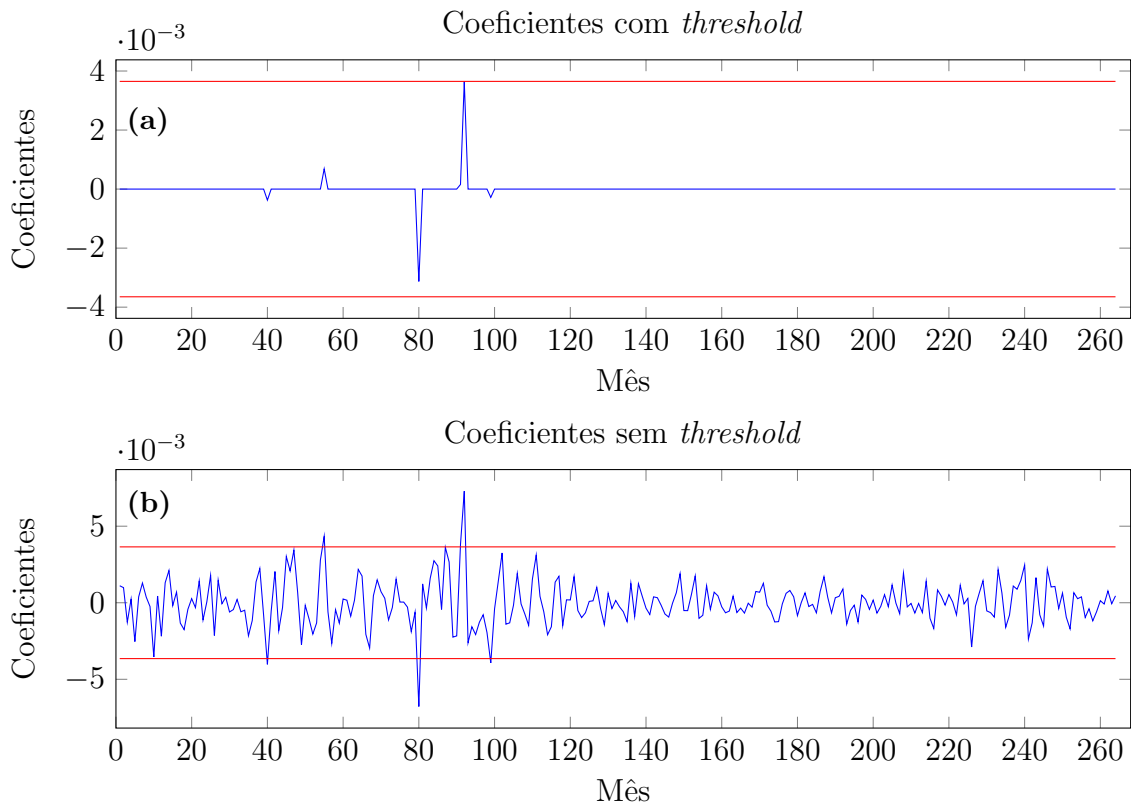
Como foi visto, as medidas de núcleo baseadas em wavelets passam no teste de viés e no teste de ajuste dinâmico, que são os testes mais usuais na literatura sobre núcleo de inflação, com evidências de que elas funcionam tão bem, ou melhor, quanto as duas medidas do Banco Central, resultado também observado por Denardin, Kozakevicius e Schmidt [16]. Na próxima seção será utilizado um método baseado em wavelets para avaliar uma possibilidade de melhorar as regressões.

#### 4.4 Medidas baseadas em wavelets com amostra restrita.

Nesta seção, começamos com uma análise dos coeficientes da Transformada Wavelet obtidos, com e sem *thresholding* e verificamos se o ajuste pode ser melhorado. A figura 4.7 mostra os coeficientes do primeiro nível da MODWT aplicados à inflação, com e sem *thresholding*. As duas linhas paralelas no eixo  $x$  mostram qual o nível de *threshold*, abaixo

do qual, em termos absolutos, os coeficientes são eliminados, e acima, são reduzidos, de acordo com a fórmula do *soft thresholding* apresentada na seção 3.3.1. Como se pode ver, todos os coeficientes em 4.7b que estão acima da linha são reduzidos em 4.7a enquanto que os que estão abaixo são zerados.

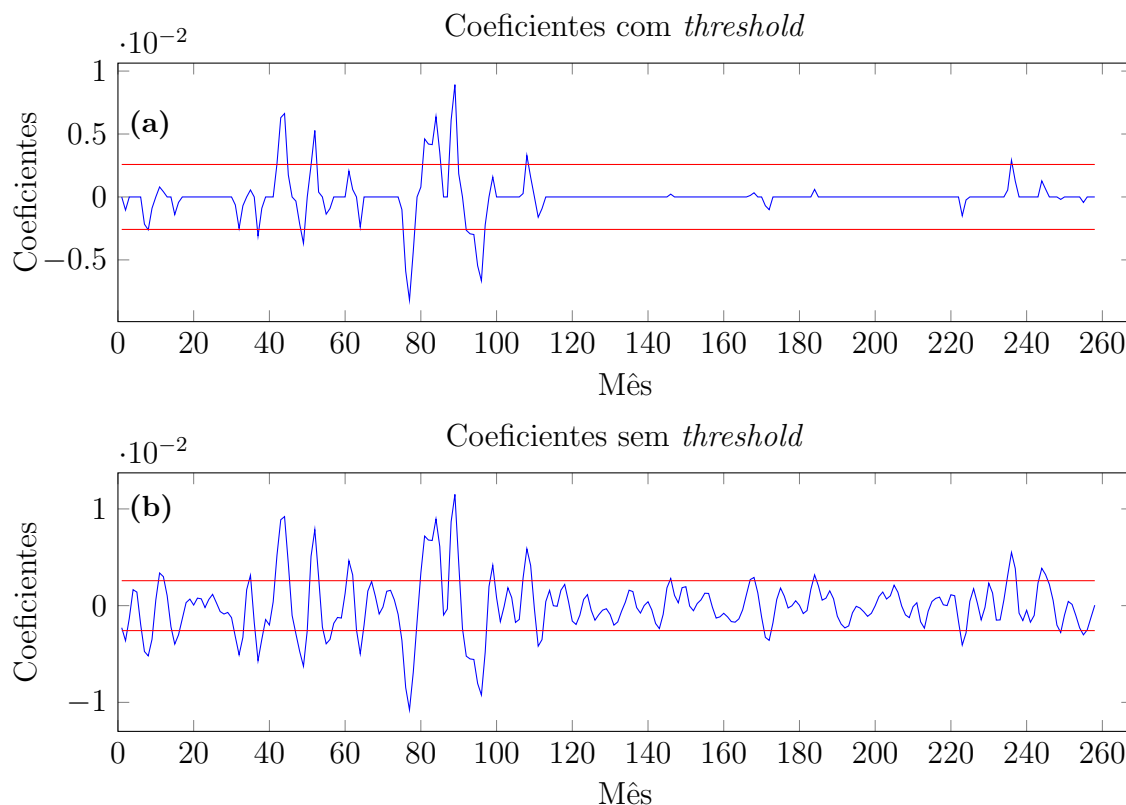
**Figura 4.7:** Coeficientes do primeiro nível *detail* da transformada wavelet



Fonte: elaborado pelo autor

A Figura 4.8 apresenta o mesmo gráfico, mas com o *threshold* e os coeficientes *detail* do segundo nível. Em ambos os gráficos parece haver um padrão: a maior parte dos coeficientes não zerados está no primeiro terço do gráfico enquanto que o segundo terço tem nenhum ou poucos coeficientes mantidos. Além disso, a redução dos coeficientes não zerados parece ser bem menor após as primeiras 120 observações. É possível que isso seja um indício de alguma forma de quebra estrutural da inflação, o que é um problema. O *threshold* requer que a variância do ruído observado junto com a inflação seja a mesma para todo o sinal. Se isso não acontecer, a metodologia não vai funcionar direito, e, como pode ser o caso, ela sobrestima o *threshold* e elimina coeficientes que deveriam ter sido mantidos.

**Figura 4.8:** Coeficientes do segundo nível *detail* da transformada wavelet



Fonte: elaborado pelo autor

Existe uma forma de se verificar se houve uma quebra estrutural utilizando a MODWT, desenvolvida por Wichter et al. [60], sendo capaz de dizer em qual observação é provável que a quebra tenha ocorrido. Aqui omitimos o desenvolvimento. Os cálculos para testar se houve uma quebra estrutural são um pouco complicados e fogem do tema deste trabalho, e portanto não serão aplicados. Em vez disso supomos que realmente houve uma quebra estrutural e aplicamos o algoritmo, como em Ramazan, Selçuk e Wichter [49], aos coeficientes *detail* do primeiro nível da transformada wavelet do IPCA.

A estatística obtida pelo teste de posição da quebra de estrutura é 115. Para verificar a possibilidade uma quebra estrutural após a observação  $n^o$  115 aplicamos um teste F para avaliar se as amostras formadas pelas observações de 1 a 114 (do IPCA) tem variância diferente das observações de 115 a 268. O resultado é que com  $p$  – valor menor que  $10^{-15}$  rejeita-se a hipótese nula de que as variâncias são as mesmas. A razão entre as variâncias dos dois períodos é 6.25. Interpretamos isso como uma quebra estrutural (na

variância do sinal) e procedemos com a estimação do núcleo de inflação utilizando apenas as observações a partir de junho de 2005 (observação n<sup>o</sup> 115).

O fato de a inflação mudar de estrutura por volta desta data foi observado por Ferreira, de Mattos e Ardeo [27], que fizeram teste de raiz unitária em duas amostras da inflação. A primeira desde janeiro de 1999 até março de 2016, para a qual, conforme foi visto na seção anterior (com dados desde dezembro de 1995), foi rejeitada a hipótese de raiz unitária para a inflação e para os núcleos oficiais, enquanto que para a amostra dos últimos 10 anos (abril de 2006 até março de 2016) foi aceita a hipótese. Isso traz evidências de que em algum momento na década de 2000 a inflação teve uma mudança estrutural.

Para esta parte do trabalho estimamos outra medida de núcleo baseada em wavelets, a por componentes desagregados, e também incluiremos a medida por médias aparadas (MA) do Banco central. A razão de inclui-las apenas aqui é que elas só podem ser calculadas, ou só existem dados oficiais para elas, a partir de 1999.

As medidas de núcleo por componentes desagregados foram propostas por Andersson [1], que em vez de aplicar a MODWT diretamente no índice de inflação cheio, aplica ela componente a componente, dado um nível de desagregação. Em vez de supor um modelo do tipo  $\pi_t = \pi_t^c + \epsilon_t$ , onde a inflação é um componente relativo ao sinal (núcleo) mais um ruído, o modelo inflacionário é dado por:

$$\pi = \pi_t^M + \sum_{k=0}^I b_k(r_{kt} + \epsilon_{kt}) \quad (4.8)$$

Onde  $\pi_t^M$  é a inflação monetária no tempo  $t$  e  $r_{kt}$  e  $\epsilon_{kt}$  são componentes idiosincráticos de cada componente do índice e  $b_k$  são os pesos da ponderação. Os termos  $\epsilon_{kt}$  são erros de medição, com esperança 0 para todo  $K$ , e os termos  $r_{kt}$  representam os choques reais que afetam o preço de cada item  $k$ , por exemplo secas, problemas de transporte etc. Sob certas condições, no longo prazo  $\sum_{k=0}^I b_k r_{kt} = 0$ , mas  $\sum_{k=0}^I r_{kt} \neq 0$  já que as pessoas podem ajustar seu consumo de forma a compensar variações nos preços relativos [1], embora não haja razão para eles se equilibrarem a cada momento. No curto prazo porém a relação pode não ser verdade, pois é necessário o ajuste. O modelo em (4.8) é semelhante ao modelo inflacionário proposto por Carl Menger [23]. Neste último,



a inflação que é causada pela administração monetária é exclusivamente dada por  $\pi_t^M$ , na terminologia de Menger, são mudanças no valor interior do dinheiro, i.e. mudanças no valor idiossincrático dele, por exemplo uma desvalorização/valorização monetária causada pela emissão/destruição de moeda. Fase e Folkertsma [23] identificam este termo como o núcleo de inflação. A preocupação do banco central de um país é assegurar que esta parte da inflação seja controlada, pois não tem controle sobre as outras<sup>16</sup> e no longo prazo as pessoas sempre podem ajustar seu consumo para acomodar os choques reais.

Segundo Andersson [1] um maneira de medir a inflação monetária é descobrir a parte da inflação que não afeta o produto no longo prazo. Quah e Vahey [48] e Lahura e Vega [40] propõem um método de isolar esta parte da inflação utilizando vetores auto regressivos. Seguindo este raciocínio, Andersson [1] propõe fazer o mesmo com as wavelets. A ideia é a mesma que aplicar a MODWT ao índice cheio. Seja  $\pi_{kt}$  a variação no preço do componente  $k$  do IPCA na data  $t$ , então

$$\pi_{kt} = \pi_t^M + r_{kt} + x_{kt} \quad (4.9)$$

Andersson [1], fazendo o mesmo raciocínio usado para supostamente separar o núcleo de inflação da inflação cheia com a MODWT, utiliza nos itens que a compõe para isolar em cada um deles o que é um choque de curto prazo do que é devido a inflação monetária e mudanças persistentes (de longo prazo) nos preços relativos na equação (4.9). Por exemplo, supõe-se que  $r_{kt} = r_{1kt} + r_{2kt}$ , onde o subscrito 1 denota a parte de longo prazo do choque e 2 a parte de curto prazo. Se esta tiver as mesmas propriedades hipotéticas do termo  $\epsilon$  em (4.2) então a MODWT vai ser capaz de expugar este efeito do preço de cada item.

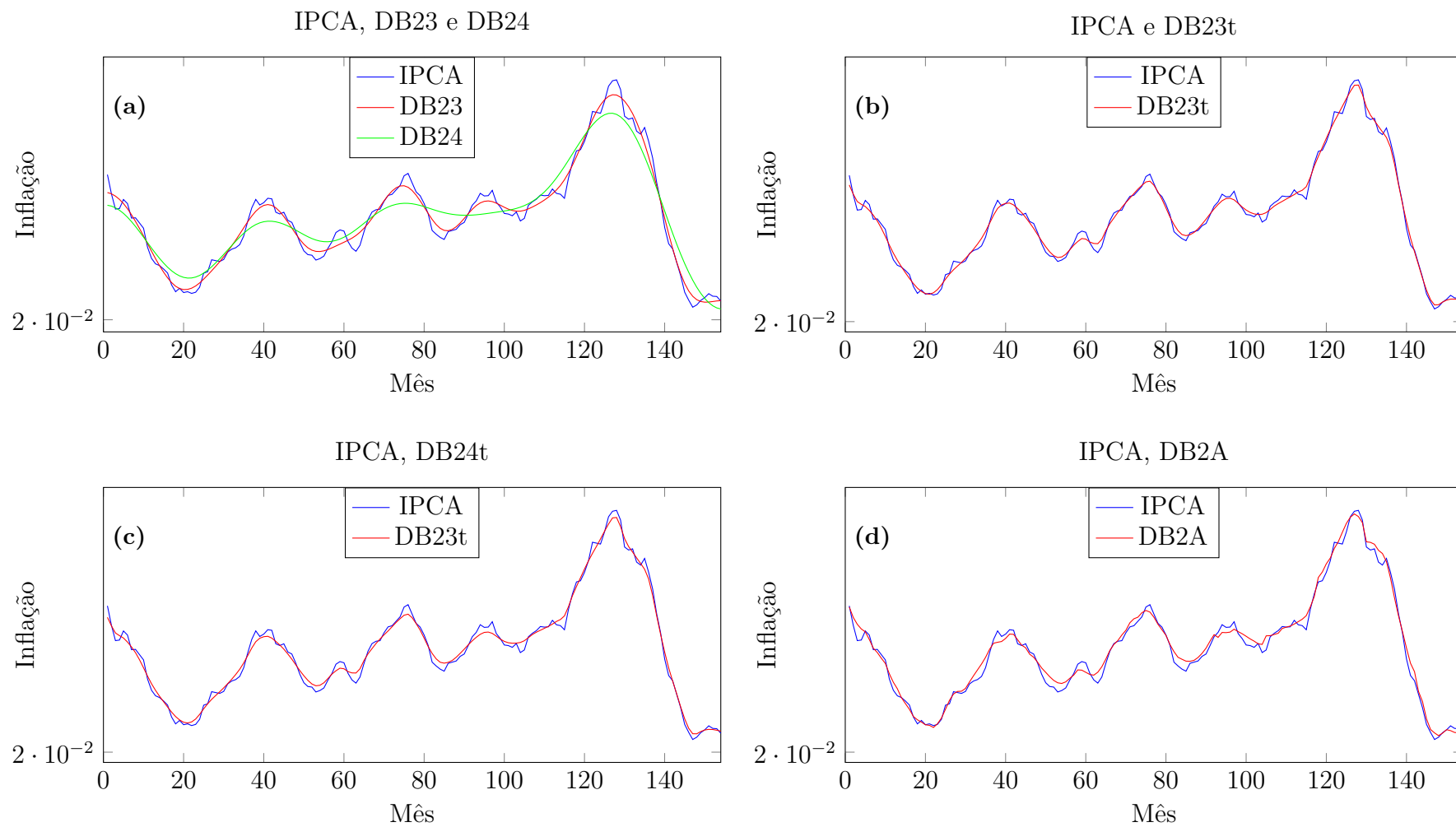
Isso é possível, da forma que será descrita adiante, por que o processo de *thresholding* não é linear e executá-lo no nível desagregado do IPCA retorna resultados diferentes. Se fossem apenas removidos níveis da MODWT de cada componente, contanto que fosse sempre removidos os mesmos níveis em todos componentes, os resultados seriam os mesmos de se aplicar a MODWT no índice cheio. Em síntese, esse método supõe que a remoção de ruídos é mais eficiente no nível desagregado.

---

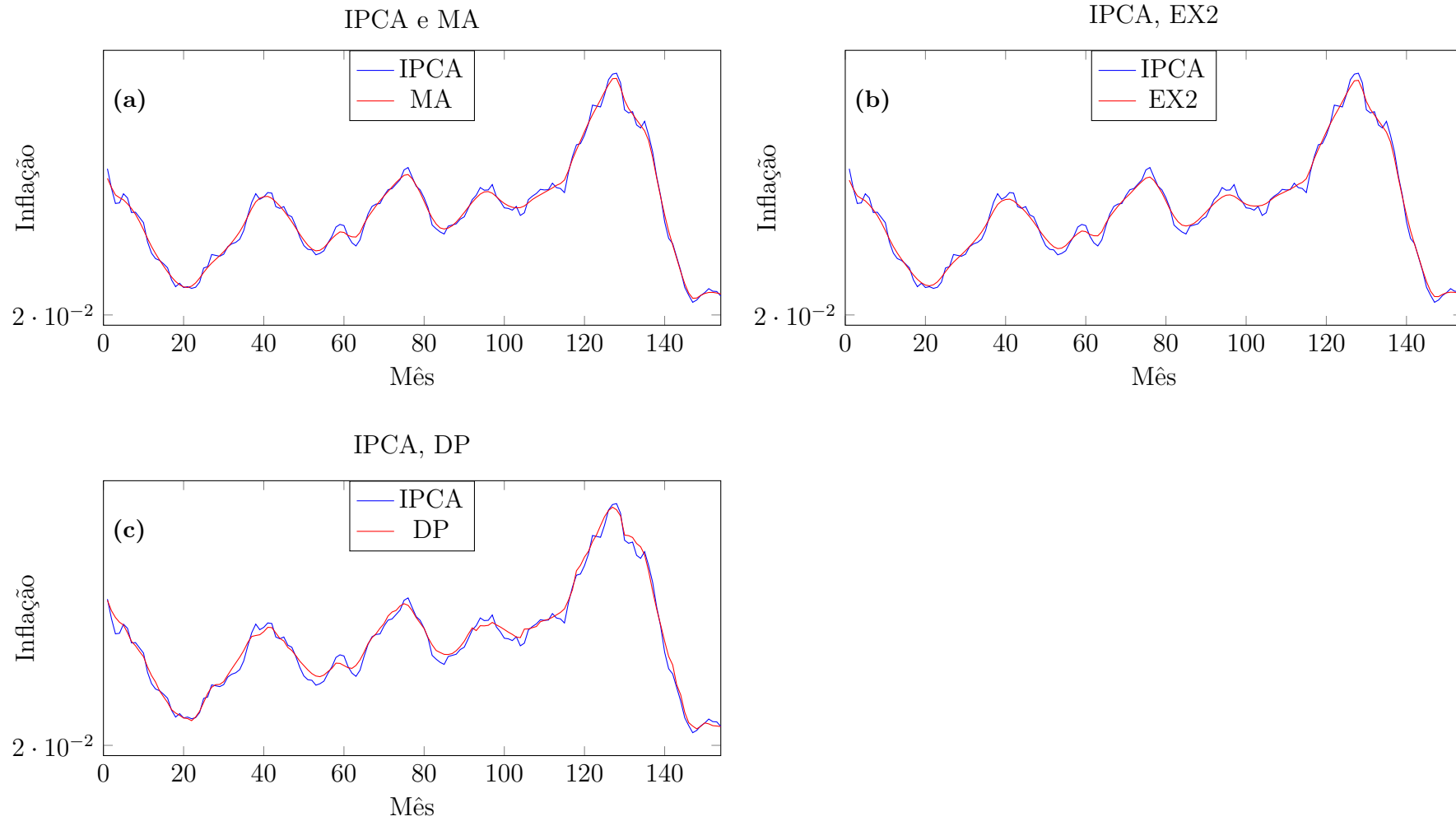
<sup>16</sup>O que não quer dizer que outros órgãos da administração governamental não influenciem os preços de outras maneiras.

Neste trabalho será apresentado um núcleo deste tipo. O núcleo será estimado com a MODWT da DB2 com *thresholding* até o terceiro nível para a desagregação do IPCA por itens. Ela será denotada por DB2A. Um problema que deve ser tratado é que os dados desagregados são mensais e os pesos mudam de mês para mês. Assim em vez de calcular a transformada para a série anualizada Andersson [1] realiza para a série mensal de cada um dos componentes e calcula o núcleo mensal do IPCA e por fim converte a série para taxas anualizadas. O núcleo mensal é calculado da mesma forma que o IPCA, mas em vez de utilizar a variação de preços em cada grupo ou item é utilizada a medida de núcleo daquele mês. Os pesos da ponderação são os mesmos que o do IPCA, o que segundo Anderson [1] garantiria que estamos estimando a inflação monetária. Isso é contrastado com as medidas de núcleo oficiais, como a MA e a DP que alteram os pesos da ponderação. Os dados desagregados do IPCA estão disponíveis em IBGE [36]. As figuras 4.9 e 4.10 mostram os gráficos dos núcleos com amostra reduzida.

**Figura 4.9:** IPCA e núcleos WIM com amostra restrita - Núcleos baseados em wavelets



**Figura 4.10:** IPCA e núcleos WIM com amostra restrita - Núcleos oficiais



A tabela 4.6 apresenta as médias, desvios padrão e a variação total das medidas de núcleo a serem avaliadas nesta parte do trabalho. Como na seção anterior, também é apresentado o p-valor do teste de normalidade (Jarque-Bera) feito nos resíduos (IPCA - Núcleo).

**Tabela 4.6:** Estatísticas descritivas e teste de normalidade para amostra reduzida

	IPCA	DB23	DB24	DB23t	DB24t	DB2A	MA	EX2	DP
<b>Média</b>	0.0574	0.0574	0.0574	0.0574	0.0574	0.0581	0.0498	0.0571	0.0587
<b>D.P.</b>	0.0184	0.0179	0.0159	0.0181	0.0178	0.0184	0.0137	0.0152	0.0143
<b>V.Total</b>	0.3767	0.2559	0.1808	0.2904	0.2746	0.2925	0.2524	0.3073	0.2817
<b>Normalidade</b>	-	0.6598	0.2308	0.503	0.4081	0.2007	0.0003	0.6577	0.0000

Fonte: elaborado pelo autor, com base nos dados do Banco Central do Brasil [4] e do IBGE [36]

Como no caso da amostra completa, todas as medidas tem a média próxima da média do IPCA, e, com a exceção da DB2A, todas as medidas baseadas em wavelets tem a mesma média. A DB2A por sua vez tem a média 7 centésimos maior. Ela também é a única medida que não tem desvio padrão amostral menor do que o desvio padrão IPCA cheio, com 0.0184. Novamente, todas as medidas de núcleo possuem variação total menor do que o IPCA. Dessa vez, o ruído extraído de todas as medidas de núcleo baseadas em wavelets passam no teste de normalidade o que indica um processo de remoção de ruídos adequado. O teste de estacionariedade é refeito na amostra reduzida. Os p-valores dos testes ADF, realizados com a função `adfTest`<sup>17</sup> [62], no nível, nas primeiras diferenças e nos ruídos são apresentados na tabela 4.7. Os testes se dão sem intercepto nem tendência.

**Tabela 4.7:** Testes de estacionariedade

	IPCA	DB23	DB24	DB23t	DB24t	DB2A	MA	EX2	DP
Nível	0.3272	1.4236	0.0100*	0.1938	0.2072	0.3144	0.3022	0.2256	0.2937
1ª Diferença	0.0100*	0.0100*	0.0100*	0.0100*	0.0100*	0.0100*	0.0100*	0.0100*	0.0100*
IPCA - Núcleo	0.000	0.0100*	0.0100*	0.0100*	0.0100*	0.0100*	0.0875	0.0100*	0.0100*

Fonte: elaborado pelos autores. com base nos dados do Banco Central do Brasil [4] e do IBGE [36]. O asterisco \* indica menor do que 1%

<sup>17</sup>Package: `fUnitRoots`, parâmetros: `x = série` e `type = "nc"`

Como se pode ver, não se pode rejeitar a hipótese de raiz unitária no nível em nenhuma das séries, exceto na DB24. Já para a segunda diferença das séries, a hipótese de não estacionariedade é rejeitada em todos os casos com p-valor menor do que 1%. Esse fato é mostrado por Ferreira, de Mattos e Arder [27] para a inflação entre abril/2006 e dez/2016. O motivo de testar a estacionariedade dos ruídos, é que segundo Wooldridge [61], se for conhecido o parâmetro potencial  $\gamma$  de cointegração entre as duas séries<sup>18</sup> então verificar se  $\pi_t - \gamma\pi_t^c$  é estacionário equivale a testar a cointegração das séries. Normalmente este parâmetro não é conhecido, mas neste caso, temos a hipótese de que  $\gamma = 1$  pois o núcleo deve ser uma medida não enviesada da inflação. A linha 3 da tabela 4.7 mostra que todas as medidas menos a MA mostram indícios de cointegração, pois rejeitam a hipótese nula de não estacionariedade, e mesmo a MA seria rejeitada com 10% de confiança. Já que ela é uma medida oficial, continuamos a fazer testes com ela.

O primeiro deles, assim como na seção anterior é o teste de viés para o núcleo. Como as séries são cointegradas podemos estimar a equação (4.5) e ter os parâmetros estimados consistentemente [61], porém os erros desta regressão tendem a ser correlacionados, e neste caso existem evidências de que eles são. Isso não é novidade pois encontrou-se o mesmo problema nos testes de viés com a amostra completa. Assim procedemos da mesma maneira. A tabela 4.8 mostra os resultados para o teste de viés. O teste F é o de Gujarati e Porter [29, p. 263] e As regressões são feitas com a função `lm`<sup>19</sup> do R.

**Tabela 4.8:** Teste de viés para amostra reduzida

	<b>DB23</b>	<b>DB24</b>	<b>DB23t</b>	<b>DB24t</b>	<b>DB2A</b>	<b>MA</b>	<b>EX2</b>	<b>DP</b>
$\alpha$	-0.0011	-0.0067	-0.0007	-0.0017	-0.0005	-0.0058	-0.0082	-0.0155
<b>D.P.</b> $\alpha$	0.0007	0.0015	0.0004	0.0005	0.0005	0.0018	0.0019	0.0017
<b>p-valor</b>	0.1119	0.0000	0.0885	0.0006	0.3979	0.0018	0.0000	0.0000
$\beta$	1.0196	1.1161	1.0128	1.0299	0.9955	1.2692	1.1475	1.2411
<b>D.P.</b> $\beta$	0.0117	0.0256	0.0071	0.0082	0.009	0.0356	0.032	0.0288
<b>p-valor</b>	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

<sup>18</sup>Duas séries,  $y_t$  e  $x_t$ , não estacionárias, são ditas cointegradas se existe uma constante  $\gamma$  tal que  $y_t - \gamma x_t$  é um processo  $I(0)$  [61, p. 719]

<sup>19</sup>Package: stats, parâmetros: IPCA  $\sim$  Núcleo

<b>Teste F</b>	0.0880	0.0000	0.0621	0.0000	0.0000	0.0000	0.0000	0.0000
----------------	--------	--------	--------	--------	--------	--------	--------	--------

Fonte: elaborado pelo autor, com base nos dados do Banco Central do Brasil [4] e do IBGE [36].

Como na amostra completa, apenas duas medidas passam no teste F (embora agora ambas passam com 5% de confiança), a DB23 com p-valor 0.0880 e a DB23t com p-valor 0.0621. Novamente, refazemos o teste com correção da covariância pelo método de Newey-West e realizamos o teste de Wald para a restrição  $\alpha = 0$  e  $\beta = 1$ . A tabela 4.9 reporta os p-valores do teste de Wald, onde a hipótese é que a restrição é verdadeira.

**Tabela 4.9:** Teste de Wald para restrição nos coeficientes com amostra restrita

	DB23	DB24	DB23t	DB24t	DB2A	MA	EX2	DP
Teste de Wald	0.4982	0.0399	0.1620	0.0017	0.0079	0.0000	0.1086	0.0000

Fonte: elaborado pelos autores. com base nos dados do Banco Central do Brasil [4] e do IBGE [36].

Neste caso, a correção da variância permite aceitar que mais duas medidas de núcleo atendem às restrições, a DB24 com p-valor de 0.0399 e a EX2 com p-valor de 0.1086. As outras medidas falham no teste. O próximo teste é o teste do ajuste dinâmico. Ele é feito da mesma forma que o teste na amostra completa, isto é, modelando a correlação dos resíduos com modelos ARIMA utilizando a função `auto.arima`<sup>20</sup>. A tabela 4.10 mostra os resultados.

**Tabela 4.10:** Teste de ajuste dinâmico com amostra restrita

	DB23	DB24	DB23t	DB24t	DB2A	MA	EX2	DP
<b>h = 6</b>								
$\beta$	-0.9597	-0.95	-0.9418	-0.9477	-0.8528	-0.9325	-0.5028	-1.0762
<b>D.P.</b> $\beta$	0.0917	0.0837	0.0989	0.0986	0.1197	0.1337	0.1336	0.1577
<b>ARIMA</b>	(2, 0, 0)	(1, 1, 0)	(2, 0, 0)	(2, 0, 0)	(2, 0, 0)	(2, 0, 0)	(2, 0, 0)	(2, 0, 1)
<b>h = 12</b>								
$\beta$	-1.2914	-1.3011	-1.3945	-1.3932	-1.3838	-1.5436	-0.6483	-1.6685
<b>D.P.</b> $\beta$	0.0829	0.0812	0.0955	0.0951	0.1202	0.149	0.1762	0.1626
<b>ARIMA</b>	(1, 1, 0)	(1, 1, 0)	(1, 1, 0)	(1, 1, 0)	(1, 1, 0)	(2, 0, 2)	(2, 0, 3)	(2, 0, 1)

<sup>20</sup>Package forecast, parâmetros: y = IPCA, xreg = núcleo.

	<b>h = 18</b>							
$\beta$	-1.1901	-1.1854	-1.3314	-1.328	-1.2979	-1.552	-0.661	-1.5656
<b>D.P.</b> $\beta$	0.0868	0.0855	0.0939	0.0935	0.1193	0.1411	0.1815	0.1684
<b>ARIMA</b>	(1, 1, 0)	(1, 1, 0)	(1, 1, 0)	(1, 1, 0)	(0, 1, 2)	(2, 0, 0)	(1, 0, 4)	(1, 0, 4)

Fonte: elaborado pelos autores. com base nos dados do Banco Central do Brasil [4] e do IBGE [36].

Para  $h = 6$  o beta de todas as medidas, exceto a EX2, é próximo de  $-1$ . E com exceção da DB24, o computador estimou para todas as medidas erros ARMA convencionais. Quando  $h$  aumenta um padrão aparece nas medidas baseadas em wavelets. A função `auto.arima` estima para todas elas ruídos da primeira etapa do GLS não estacionários e toma suas primeiras diferenças (o termo do meio nas triplas da linha ARIMA é a ordem de diferenciação). Isso pode ocorrer por que o ruído resultante é não estacionário, ou por causa de algum fator sazonal. Aqui não investigamos isso a fundo, e como os resultados estão de acordo com o esperado, e de acordo com os resultados feitos na amostra cheia, aceitamos os valores da tabela 4.10. Também notamos que as séries dependentes mostram indícios de não estacionariedade para  $h = 12$  e  $18$ , porém segundo Wooldridge [61], o GLS lida com a não estacionariedade da séries, pelo menos aproximadamente.

O último teste feito para avaliar as medidas de núcleo é o teste de poder de predição da inflação futura. Ele é feito da mesma forma que na seção anterior. Guardamos um terço da amostra, estimamos modelos com a função `auto.arima`<sup>21</sup>, fazemos previsões 12 passos a frente para cada núcleo com a função `forecast`<sup>22</sup> e geramos uma estatística de raiz do erro quadrado médio com a função `accuracy`<sup>23</sup>. Em seguida a amostra é acrescida de uma observação e o teste é refeito, e assim procedemos até que toda a amostra, menos as 12 últimas observações sejam utilizadas para modelagem. A tabela 4.11 apresenta os resultados. A coluna à direita apresenta a razão entre a média das RMSE's dos modelos de previsão com o núcleo para os modelos *benchmark* que usam a *lags* da inflação.

<sup>21</sup>Package `forecast`, parâmetros:  $y = \text{IPCA}$ ,  $x_{\text{reg}} = \text{núcleo}$ .

<sup>22</sup>Package `forecast`, parâmetros:  $object = \text{regressão}$ ,  $h = 12$ .

<sup>23</sup>Package `forecast`, parâmetros:  $f = \text{objeto } forecast$ ,  $x = \text{IPCA}$ .



**Tabela 4.11:** Teste de poder preditivo dos núcleos

	Média RMSE	Razão <i>Benchmark</i>
<b>IPCA</b>	0.0203	1.000
<b>DB23</b>	0.0269	1.3278
<b>DB24</b>	0.0266	1.3098
<b>DB23t</b>	0.0198	0.9762
<b>DB24t</b>	0.0195	0.9631
<b>DB2A</b>	0.0193	0.9491
<b>MA</b>	0.0185	0.9122
<b>EX2</b>	0.0160	0.7879
<b>DP</b>	0.0184	0.9087

Fonte: elaborado pelos autores, com base nos dados do Banco Central do Brasil [4] e do IBGE [36].

Diferente da amostra completa, as previsões feitas com os núcleo agora são em média mais eficientes, com exceção das medidas baseadas em wavelets com remoção de níveis sem *thresholding*. A medida de núcleo que melhor se saiu prevendo a inflação futura foi a EX2, com um índice de erro de quase 79% do *benchmark*. Notamos porém que estes testes são bem restritos, pois só são utilizados modelos ARIMA em cada modelo, sem modelos de defasagens distribuídas, por exemplo. Uma previsão mais acurada do IPCA provavelmente incluiria outras variáveis, como taxa de juros câmbio, crescimento econômico, etc. Por outro lado, modelos simples tem suas vantagens, por exemplo, menos interação entre variáveis explicativas.

A última análise que mostramos é um comparação dos RMSE's da amostra cheia e da amostra restrita. A tabela 4.12 mostra a média dos últimos 40 RMSE's obtidos com a amostra cheia, lembrando que nem a MA nem a DB2A foram calculadas. Como ela só inclui as últimas 40 observações, isto é, o número de previsões feito com a amostra restrita, podemos comparar com a tabela 4.11 para saber qual estratégia é melhor, estimar com a amostra completa ou com a amostra restrita. Todas as medidas que não são baseadas em wavelets se saíram melhor, ou muito próximas no caso da EX2, com a amostra cheia do que com a amostra restrita, para os modelos selecionados pela *auto.arima*. Em particular,

nos modelos com o IPCA os últimos RMSE's tem média 0.0176 com dados da amostra cheia, e 0.0203 com dados da amostra restrita, um aumento de 15%. O aumento observado para a EX2 é de 6%.

**Tabela 4.12:** Comparação de RMSE

	Média RMSE	Razão <i>Benchmark</i>
<b>IPCA</b>	0.0176	1.000
<b>DB23</b>	0.0309	1.7546
<b>DB24</b>	0.0280	1.5873
<b>DB23t</b>	0.0201	1.1401
<b>DB24t</b>	0.0202	1.1447
<b>EX2</b>	0.0150	0.8501
<b>DP</b>	0.0188	1.0665

Fonte: elaborado pelo autor, com base nos dados do Banco Central do Brasil [4]

Já as medidas baseadas em wavelets, todas elas foram ligeiramente mais eficientes prevendo quando utilizada a amostra restrita. O destaque fica com a DB23 que teve um erro médio 12.8% menor com a amostra restrita. Lembrando que a amostra foi reduzida para tentar usar o algoritmo de *thresholding* de forma mais eficiente, parece haver motivos para tal, pelo menos do ponto de vista preditivo quando o uso de núcleos baseados em wavelets. Por outro lado, a predição feita de outras maneiras parece ser mais eficiente de qualquer jeito, o que no fim das contas, não justifica restringir a amostra. É importante notar que não foi feito o procedimento padrão de Witcher et. al. [60] para verificar se houve mesmo uma mudança de variância nas séries.

Como nos outros trabalhos que envolvem a utilização de wavelets (referências [1], [2], [16], [18], [40] e [47]) foi mostrado que núcleos de inflação baseados em wavelets satisfazem os testes desenhados para verificar se uma dada medida de núcleo é adequada. A comparação mais importante é com o trabalho de Denardin, Kozakevicius e Schmidt, feito para o caso brasileiro. As diferenças do seu trabalho para este é que aqui utilizamos a etapa de *thresholding* para estimar algumas medidas e que fazemos também a MRA até o nível 4, em vez de apenas até o nível 3.

## 5 CONCLUSÃO

Neste trabalho foi feita uma revisão teórica das wavelets e de algumas de suas propriedades, e então foi apresentada uma aplicação da ferramenta ao problema econômico da estimação do núcleo de inflação.

Foi visto que as funções sinusoides, quando amostradas de uma forma particular, formam uma base para  $\mathbb{R}^N$ , chamada Base de Fourier. Os vetores desta base são identificados como frequências. Nos casos que vimos, os vetores que procuramos expressar nesta base eram sinais, ou amostras de sinais, que por serem observados em quantidade finita em em tempos discretos, são bem representados por vetores em espaços de dimensão finita.

A Base de de Fourier tem algumas propriedades importantes. Ela é composta por vetores ortogonais, o que faz a obtenção dos coeficientes que expressam qualquer  $x$  nela mais fácil. A obtenção deles através do cálculo de produtos internos é chamada de Transformada Discreta de Fourier (DFT) . Os coeficientes desta representação são denotados  $\mathbf{X}_k$  e seu valor absoluto é chamado de espectro. É possível também a aplicação do Teorema de Parseval, que diz que a energia de um vetor é dada pela soma dos quadrados dos coeficientes de sua representação em uma base ortogonal. A aplicação do Teorema de Parseval, permite justificar a remoção de ruídos de um sinal. Mas para este trabalho, a propriedade mais utilizada da base de Fourier é a sua ligação com matrizes circulares. Como foi apresentado, as matrizes circulares são a representação da transformada linear chamada convolução circular, ou apenas convolução, denotada  $(x * w)$ . Ela é formada empilhando-se  $N$  rotações de um vetor linha até obter uma matriz  $N \times N$ .

A convolução entre dois vetores resulta em um novo vetor que possui espectro, entrada a entrada, igual ao produto dos espectros dos dois vetores. Para aplicar a convolução são utilizados vetores com espectro previamente conhecido, chamados filtros. Com filtros, agimos no espectro de um vetor sem calcular sua DFT. Se o filtro for um filtro passa-baixo, isto é, um filtro com DFT pequena para as frequências mais altas, então sua convolução com um sinal vai atenuar as frequências mais altas. Um filtro passa-alto faz o mesmo com as frequências mais baixas. E entre os filtros possíveis, estão so chamados

Filtros Wavelets, que sempre vem em conjunto, um filtro passa-baixo, chamado função escala, e um filtro passa-alto, chamado wavelet.

Foi apresentado como os Filtros Wavelet após operações de sub amostragem (decimamento) dão origem a uma nova transformada ortogonal, chamada Transformada Wavelet Discreta (DWT), que pode ter diversos níveis, indexados por uma escala. A expansão de um vetor nas colunas da transposta desta matriz é chamada de Análise Multirresolução (MRA). Além de representar o sinal no domínio da frequência, a MRA tem resolução temporal, embora as duas não possam ser vistas simultaneamente neste trabalho. Para aplicações a series temporais a DWT sofre de uma deficiência importante: ela não é capaz de lidar com rotações de um sinal, o que implica que ela retorna resultados diferentes dependendo da data inicial das observações.

Uma das formas de contornar isso é a utilização dos Filtros Wavelet sem decimamento e uma das formas de se fazê-lo é com a *Maximum Overlap Discrete Wavelet Transform* (MODWT), que pode ser tanto vista como uma aplicação de dois filtros (um passa-alto e um passa-baixo) em conjunto, ou como uma aplicação dupla da DWT a um sinal e a uma rotação dele. A vantagem de ver a MODWT como um filtro é que podemos utilizar as ferramentas da análise de Fourier para estudar suas propriedades. A desvantagem da MODWT é que ela não é uma transformada ortogonal. A MODWT também gera uma MRA com a qual podemos representar um sinal em diferentes escalas.

Além de ser um conjunto ortogonal, as colunas da DWT formam uma base esparsa de  $\mathbb{R}^N$ , o que significa que vetores do espaço, que atendam a certos requisitos dependendo da wavelet utilizada, podem ser reescritos, com pequena perda de informação, com poucos componentes da base. Se um sinal for observado em conjunto com algum ruído, essa propriedade permite separar os valores do sinal dos valores do ruído. O uso da esparsidade desta forma é operacionalizado no algoritmo de *thresholding* que seleciona e redimensiona os coeficientes que devem ser utilizados na reconstrução do sinal para que este seja o mais próximo do sinal verdadeiro sem ruídos. Este procedimento é apresentado para DWT, e a extensão para MODWT é imediata, notando a ligação entre as duas.

Na segunda parte do trabalho, as ferramentas que foram apresentadas, especialmente a MODWT, são aplicadas ao problema da estimação do núcleo de inflação.

O núcleo de inflação é uma ferramenta disponível para análise de um índice de preços. Quando observado diretamente, o índice de preços pode apresentar ruídos que dificultam a avaliação da situação inflacionária em uma economia. Por exemplo, choques que não tem nenhuma correlação com a política monetária podem afetar preços, para cima ou para baixo, sem que seja necessária uma interferência do banco central. Estes choques são passageiros, então idealmente, não deveriam influenciar a política monetária. O núcleo de inflação é entendido como a parte da inflação que não sofre efeito destes choques e pode ser utilizado como referencia. O problema é como estimá-lo.

Diversas formas de estimação são sugeridas pela literatura, como remoção de componentes (núcleos por exclusão), remoção dos componentes com maior oscilação e também modelos de estimação de sinais, entre eles, a MODWT. Neste trabalho, apresentamos núcleos de inflação construídos com wavelets e comparamos eles com núcleos oficiais calculados pelo Banco Central do Brasil. Para avaliar estes núcleos são feitos testes sugeridos pela literatura. Eles são: teste de viés, teste de ajuste dinâmico e teste de poder de previsão.

O teste de viés serve para verificar se uma medida de núcleo está mesmo estimando algo que se pode chamar de núcleo de inflação. O teste de ajuste dinâmico verifica se desvios do núcleo tendem a ser corrigidos através do tempo, isto é, se a inflação está acima/abaixo dele, então ela deve cair/subir. O teste de poder preditivo verifica se alguém estaria melhor se predissesse a inflação com núcleo em vez de utilizar a própria inflação.

Nos dois primeiros testes, as medidas de núcleo baseadas em wavelets, em geral se saem bem. Há evidências que elas são pelo menos tão boas quanto as medidas oficiais para medir a inflação subjacente e o teste de ajuste dinâmico mostra que a diferença entre inflação e núcleo baseado em wavelets pode dar pistas para a tendência inflacionária. Quanto ao teste preditivo, não foi mostrado que previsões feitas com o núcleo são mais eficientes para prever inflação futura.

A aplicação das wavelets nos trabalhos que estimaram o núcleo de inflação em geral é feita da forma mais tradicional de processamento de sinais, isto é, um conjunto de informação é separado em diversos componentes e reconstruído com a remoção de alguns deles. Existem outras possibilidades de executar esta tarefa, como aponta Steehouwer

[56], que sugere , modelar os termos separadamente (em vez de removê-los). Além disso como existem outras formas de seleção de coeficientes além do *soft thresholding*, como as descritas por Percival e Walden [46] no capítulo 10. Trabalhos futuros podem focar em modelar os componentes das diferentes escalas do MRA ou em como selecionar os coeficientes a serem mantidos no processo de *thresholding*, por exemplo, Smith [55] sugere, a utilização de redes neurais (*machine learning*) para escolha de parâmetros de *benchmark*.

## Referências Bibliográficas

- [1] ANDERSSON, F.N.G. **Core inflation** - why the FED got it wrong. 2008. Disponível em: <<http://portal.research.lu.se/portal/files/6135358/2061638>>. Acesso em 06/02/2018.
- [2] BAQAEE, David. Using wavelets to measure core inflation: the case for New Zealand. **North American Journal of Economics and Finance**. Elsevier, v. 21. 2010.
- [3] BANCO CENTRAL DO BRASIL. Regime de metas para a inflação no Brasil informações até junho de 2016. **Série Perguntas Mais Frequentes**. Brasília: Banco Central do Brasil, 2016.
- [4] BANCO CENTRAL DO BRASIL. **Sistema de séries temporais**. 2018. Disponível em: <<https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries>>. Acesso em 06/02/2018.
- [5] BANCO CENTRAL DO BRASIL. **Relatório de inflação**. Brasília, Banco Central do Brasil, v.11 n. 4, dez 2009. Disponível em <<http://www.bcb.gov.br/htms/relinf/port/2009/12/ri200912inp.pdf>>. Acesso em 18/04/2018.
- [6] BANCO CENTRAL DO BRASIL. **Relatório de inflação**. Brasília, Banco Central do Brasil, v.20 n. 1, mar de 2018. Disponível em <<http://www.bcb.gov.br/htms/relinf/port/2018/03/ri201803P.pdf>>. Acesso em 18/04/2018.
- [7] BLINDER, Alan S. Commentary. **Federal Reserve of St. Louis Review**. St. Louis: Federal Reserve Bank of St. Louis, May/Jun 1997.
- [8] BLINDER, Alan S.. Monetary policy today: sixteen questions about twelve answers. **CEPS Working Papers**. n. 129, Jul 2006.
- [9] BROUGHTON, S. Allen; BRYAN, Kurt. **Discrete Fourier analysis and wavelets**. Hoboken: John Wiley & Sonas, Inc., 2008.

- [10] BRYAN, Michael F.; CHECCHETTI, Stephen G.. Measuring core inflation. **Monetary Policy**. Chicago: Chicago University Press, January 1994.
- [11] BUREAU OF LABOR STATISTICS. **Handbook of methods**. Washington: Bureau of Labor Statistics, Dec. 2017. Disponível em <<https://www.bls.gov/opub/hom/pdf/homch17.pdf>>. Acesso em 18/04/2018.
- [12] BUREAU OF LABOR STATISTICS. **Consumer Price Index**. Disponível em: <<https://www.bls.gov/cpi/research-series/home.htm#CPI-U-R\%20Data>>. Acesso em 20/04/2018.
- [13] CHECCHETTI, Stephen G.. Measuring short-run inflation for central bankers. **Federal Reserve Bank of St. Louis Review**. St. louis: Federal Reserve Bank of St. Louis, v. 79, n. 3, May/Jun 1997.
- [14] CHUI, Charles K. **An introducton to wavelets**. San Diego: Academic Press, 1992.
- [15] COGLEY, Timothy. A simple adaptative measure of core inflation. **Journal of Money, Credit and Banking**. Ohio: Ohio University Press, v. 34, n. 1. Feb., 2002.
- [16] DENARDIN, Anderson A.; KOZAKEVICIUS, Alice; SCHMIDT, Alex A.. **Avaliação da medida de núcleo de inflação baseada no método wavelet para o brasil**. Niterói: Anpec, 2015. Disponível em: <[https://www.anpec.org.br/sul/2016/submissao/files\\_I/i6-cee5015ba5b83b516b3751cb09321629.pdf](https://www.anpec.org.br/sul/2016/submissao/files_I/i6-cee5015ba5b83b516b3751cb09321629.pdf)>. Acesso em 06/02/2018.
- [17] DONOHO, DAVID L.; JHONSTONE, Ian M.. Ideal Spatial Adaptation by Wavelet Shrinkage. **Biometrika**, Oxford: Oxford University Press, v. 81, n. 3, Aug. 1994, p. 425-255.
- [18] DOWD, Kevin; COTTER, John; LOH, Lixia. US core inflation: a wavelet analysis. **Macroeconomics Dynamics**. Cambridge: Cambridge University Press, v. 15, 2011.
- [19] EVIEWS. Robust standard errors. **Eviews user's guide**. Irwine: IHS Global Inc. Dec. 2017. Disponível em: <<http://www.eviews.com/help/helpintro>>.



html#page/content\%2FRegress2-Robust\_Standard\_Errors.html\%23>.  
Acesso em 18/04/2017.

- [20] EViews. Estimating ARIMA and ARFIMA Models in EViews. **Eviews user's guide**. Irwine: IHS Global Inc. Dec. 2017. Disponível em: <[http://www.eviews.com/help/helpintro.html#page/content/timeser-Estimating\\_ARIMA\\_and\\_ARFIMA\\_Models\\_in\\_EViews.html](http://www.eviews.com/help/helpintro.html#page/content/timeser-Estimating_ARIMA_and_ARFIMA_Models_in_EViews.html)>. Acesso em 18/04/2017.
- [21] EViews. Unit root testing. **Eviews user's guide**. Irwine: IHS Global Inc. Dec. 2017. Disponível em: <[http://www.eviews.com/help/helpintro.html#page/content/advtimeser-Unit\\_Root\\_Testing.html](http://www.eviews.com/help/helpintro.html#page/content/advtimeser-Unit_Root_Testing.html)>. Acesso em 18/04/2017.
- [22] EViews. Coefficient diagnostics **Eviews user's guide**. Irwine: IHS Global Inc. Dec. 2017. Disponível em: <[http://www.eviews.com/help/helpintro.html#page/content/testing-Coefficient\\_Diagnostics.html](http://www.eviews.com/help/helpintro.html#page/content/testing-Coefficient_Diagnostics.html)>. Acesso em 07/05/2018. .
- [23] FASE, M. M. G; FOLKERTSMA, C. K.. **Measuring inflation**: an attempt to operationalize carl menger's concept of the inner value of money. 1997. Disponível em: <[https://www.bis.org/publ/bisp05\\_p8.pdf](https://www.bis.org/publ/bisp05_p8.pdf)>. Acesso em 18/04/2017.
- [24] FEDERAL RESERVE OF NEW YORK. A review of core inflation and an evaluation of its measures. **Staff Reports**.New York, n. 236, Dec 2015.
- [25] FEDERAL RESERVE OF ST. LOUIS. **Federal Funds target range** - upper limit. Disponível em: <<https://fred.stlouisfed.org/series/DFEDTARU>>. Acesso em 20/04/2018.
- [26] the FEDERAL RESERVE. **Minutes of the Federal Open Market Comittee** January 30 - 31, 2018. Disponível em: <<https://www.federalreserve.gov/monetarypolicy/files/fomcminutes20180131.pdf>>. Acesso em 20/04/2018.
- [27] FERREIRA, Pedro Costa; de MATTOS, Daiane Marcolino; ARDEO, Vagner Laerte. Triple-filter core inflation: a measure of the inflation trajectory. **Revista Brasileira de Economia**. v. 71, n.4, p. 397-411, Out-Dez 2017.

- [28] FOURIER, Joseph. **Théorie analytique de la chaleur**. Sceaux: Jacques Gabay, 1988. Disponível em: <[https://www.irphe.fr/~clanet/otherpaperfile/articles/Fourier/N0029061\\_PDF\\_1\\_676.pdf](https://www.irphe.fr/~clanet/otherpaperfile/articles/Fourier/N0029061_PDF_1_676.pdf)>
- [29] GUJARATI, Damodar N.. PORTER, Dawn C.. **Econometria Básica**. Trad. Denise Durante, Mônica Rosenberg, Maria Lúcia G.L. Rosa. Porto Alegre: AMGH, 5 ed, 2011.
- [30] HAAR, Alfred. **On the theory of orthogonal function systems**. Trad. Georg Zimmermann. 2009. Disponível em: <<https://pdfs.semanticscholar.org/3b08/b61ba914626db518b6add5b73ac21d62f0c1.pdf>>. Acesso em: 22/06/2018.
- [31] HARVEY, Andrew; TRIMBUR, Thomas. Trend estimation and the Hodrick-Prescott filter. **Journal of the Japan Statistical Society**. Japan Statistical Society, Tokyo , v. 38, n. 1, 2008. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.323.9040\&rep=rep1\&type=pdf>>. Acesso em 12/03/2018.
- [32] HUNT, B.R.. A matrix theory proof of the discrete convolution theorem. **IEEE Transaction of Audio and Electroacoustics**. Institute of Electrical and Electronics Engineers, Piscataway, v. 19, n. 4, Dec, 1971.
- [33] HYNDMAN, R.; BERGMEIR, C.; CARCERES, G.; CHHAY, L.; O'HARAWILD, M. PETROPOULOS, F.; RASBASH, S.; WANG, E.; YASMEEN, F.. Forecast: forecasting functions for time series and linear models. **R package version 8.3**. Disponível em: <<http://pkg.robjhyndman.com/forecast>>. Acesso em 07/05/2018.
- [34] HYNDMAN, R. J.; KHANDAKAR, Y.. Automatic time series forecasting: the forecast package for R. **Journal of Statistical Society**. v. 26, n. 3, p. 1-21, 2008. Disponível em: <<http://www.jstatsoft.org/article/view/v027i03>>. Acesso em 07/05/2018.
- [35] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Sistema nacional de índice de preços ao consumidor métodos de cálculo**. 7. ed. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2013.

Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv65477.pdf>>. Acesso em 06/02/2018.

- [36] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Sistema de Séries Temporais**. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2018. Disponível em: <<https://sidra.ibge.gov.br/home/pimpfbr/brasil>>. Acesso em 06/02/2018.
- [37] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Índice de nacional de preços ao consumidor amplo - IPCA e índice nacional de preços ao consumidor - INPC**. Disponível em: <[https://ww2.ibge.gov.br/home/estatistica/indicadores/precos/inpc\\_ipca/defaultinpc.shtm](https://ww2.ibge.gov.br/home/estatistica/indicadores/precos/inpc_ipca/defaultinpc.shtm)>. Acesso em 06/02/2018.
- [38] JOHNSTONE, IAN. Wavelets and the theory of non-parametric function estimation. **Philosophical Transactions of The royal Society**. London, v. 357, n. 1770, p. 2475-2493, Sep 1999.
- [39] JONES, Charles I.. **Teoria do crescimento econômico**. Trad. Maria José Cyhlar Moneiro. Rio de Janeiro: Elsevier, 2000.
- [40] LAHURA, Erick; MARCO, Vega. Evaluation of wavelet-based core inflation measures: evidence from Peru. **Serie de Documentos de Trabajo**. Lima, n. 2011-019, Dic 2011.
- [41] LAX, Peter. **Linear algebra** and its applications. 2. ed. New york: Wiley, 1997.
- [42] LAY, David C.. **Álgebra linear e suas aplicações**. Trad. Ricardo Camelier e Valéria de Magalhães Iório. 2<sup>a</sup> ed. Rio de Janeiro: LTC, 2012.
- [43] MALLAT, Stephane G.. A theory for multiresolution signal decomposition: the wavelet representation. **IEEE Transactions of Pattern Analysis and Machine Intelligence**. Piscataway, v. 11, n. 17, p. 674-93, Jul 1989.
- [44] MALLAT, Stephane G.. **A wavelet tour of signal processig** the sparce way. Burlington: Elsevier, 2009.

- [45] NIST/SEMATECH. **e-Handbook of statistical methods**. Washington: U.S. Department of Commerce, 2012. Disponível em: <<http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc43.htm>>. Acesso em: 23/11/2017.
- [46] PERCIVAL, Donald B.; WALDEN, Andrew. **Wavelet methods for time series analysis**. Cambridge: Cambridge University Press, 2000.
- [47] du PLESSIS, Stan; du GIDEON, Rand; KOTZÉ, Kevin. Measuring core inflation in South Africa. **ERSA Working Paper**. n. 503, Jan 2015.
- [48] QUAH, Danny; VAHEY, Shaun. Measuring core inflation. **Economic Journal**. London, v. 105, n. 432, p. 1130-44, 1995.
- [49] RAMAZAN, Gençay; SELÇUK, Faruk; WITCHER, Brandon. **An introduction to wavelets and othe filtering methods in finance and economics**. San Diego: Academic Press, 2002.
- [50] SANTOS, Cristiano; CASTELAR, Ivan. Avaliando as medidas de núcleo de inflação no Brasil. **Economia Aplicada**. São Pualo: Universidade de São Paulo, v. 20, n. 1, 2016.
- [51] SCHLEICHER, Christoph. An introduction to wavelets for economists. **Bank of Canada working paper 2002-3**. Jan. 2003. Disponível em: <<https://www.banqueducanada.ca/wp-content/uploads/2010/02/wp02-3.pdf>>. Acesso em: 29/05/2018.
- [52] SHANNON, Claude E.. Communication in the presence of noise. **Proceedings of the IEEE**. Piscataway, v. 86, n. 2, Feb, 1998. Disponível em: <<http://nms.csail.mit.edu/spinal/shannonpaper.pdf>>. Acesso em 23/11/2017.
- [53] da SILVA FILHO, Tito Nícias Teixeira; FIGUEIREDO, Francisco Marcos Rodrigues. Has inflation core been doing a good job in brazil. **Revista Brasileira de Economia**. Rio de Janeiro, v. 65, n. 2. p. 207-233. Abr-Jun 2011. Disponível em: <<http://www.scielo.br/pdf/rbe/v65n2/a05v65n2.pdf>>. Acesso em 15/03/2018.
- [54] da SILVA FILHO, Tito Nícias Teixeira; FIGUEIREDO, Francisco Marcos Rodrigues. Revisitando as medidas de núcleo de inflação do Banco Central do Brasil.

- Trabalhos para Discussão**. Brasília: Departamento de Estudos e Pesquisas do BACEN, n. 365, maio 2014. Disponível em: <<https://www.bcb.gov.br/pec/wps/port/TD356.pdf>>. Acesso em 06/02/2018.
- [55] SMITH, Steven W. **The scientist and engineer's guide to digital signal processing**. Disponível em: <<http://www.dspguide.com/ch15/1.htm>>. Acesso em 23/11/2017.
- [56] STEEHOUWER, Hens. A frequency domain methodology for time series modeling. In: BERKELAAR, A.; COCHE, J.; NYHOLM, K. (Ed.). **Interest Rate Models, Asset Allocation and Quantitative Techniques for Central Banks and Sovereign Wealth Funds**. Basingstoke: Palgrave MacMillan, 2010. p. 280 - 324.
- [57] STRUIK, Dirk Jan. Joseph Fourier. **Encyclopædia Britannica**. Chicago: Encyclopædia Britannica, inc, Jan 2017. Disponível em <<https://www.britannica.com/biography/Joseph-Baron-Fourier>>, Acesso em: 23/11/2017
- [58] WALKER, James S. **A primer on wavelets and their scientific applications**. Boca Raton: CRC Press, 1999.
- [59] WAVELET BROWSER. **Wavelet properties browser**. Disponível em: <<http://wavelets.pybytes.com/>>. Acesso em 20/04/2018.
- [60] WICHTER, Brandon; BYERS, Simon D.; GUTTORP, Peter; PERCIVAL, Donald. Testing for homogeneity of variance in time series: long memory, wavelets and the Nile river. **Water Resources Research**. v. 38, n. 5, 12.1-12.16, May 2002. Disponível em: <<https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/2001WR000509>>. Acesso em 20/04/2018.
- [61] WOODRIDGE, Jeffrey M.. **Introdução à econometria** uma abordagem moderna. trad. Priscilla Rodrigues da Silva Lopes e Livia Marina Koepl. 6. ed. São Paulo: Cengage Learning, 2016.
- [62] WUERTZ, Diethelm; SETZ, Tobias; CHALABI, Yohan. fUnitRoots: Rmetrics - modelling trends and unit roots. **R package version 3042.79**. Dis-

ponível em: <<https://CRAN.R-project.org/package=fUnitRoots>>. Acesso em: 07/05/2018.

[63] WYNNE, Mar A. Core inflation: a review of some conceptual issues. **Federal Reserve of St. Louis Review**. St. Louis: Federal Reserve Bank of St. Louis, v.90, n.3, May/Jun 2008.

[64] HYNDMAN, Rob; KHANDAKAR, Yeasmin. Automatic time series forecasting: the forecast package for R. **Journal of Statistical Society**. v. 26, n. 3, p. 1-21, 2008. Disponível em: <<http://www.jstatsoft.org/article/view/v027i03>>. Acesso em 07/05/2018.

## 6 APÊNDICE

Figura 6.1: Tabela dos itens que compõe o núcleo por exclusão 2 (EX2)

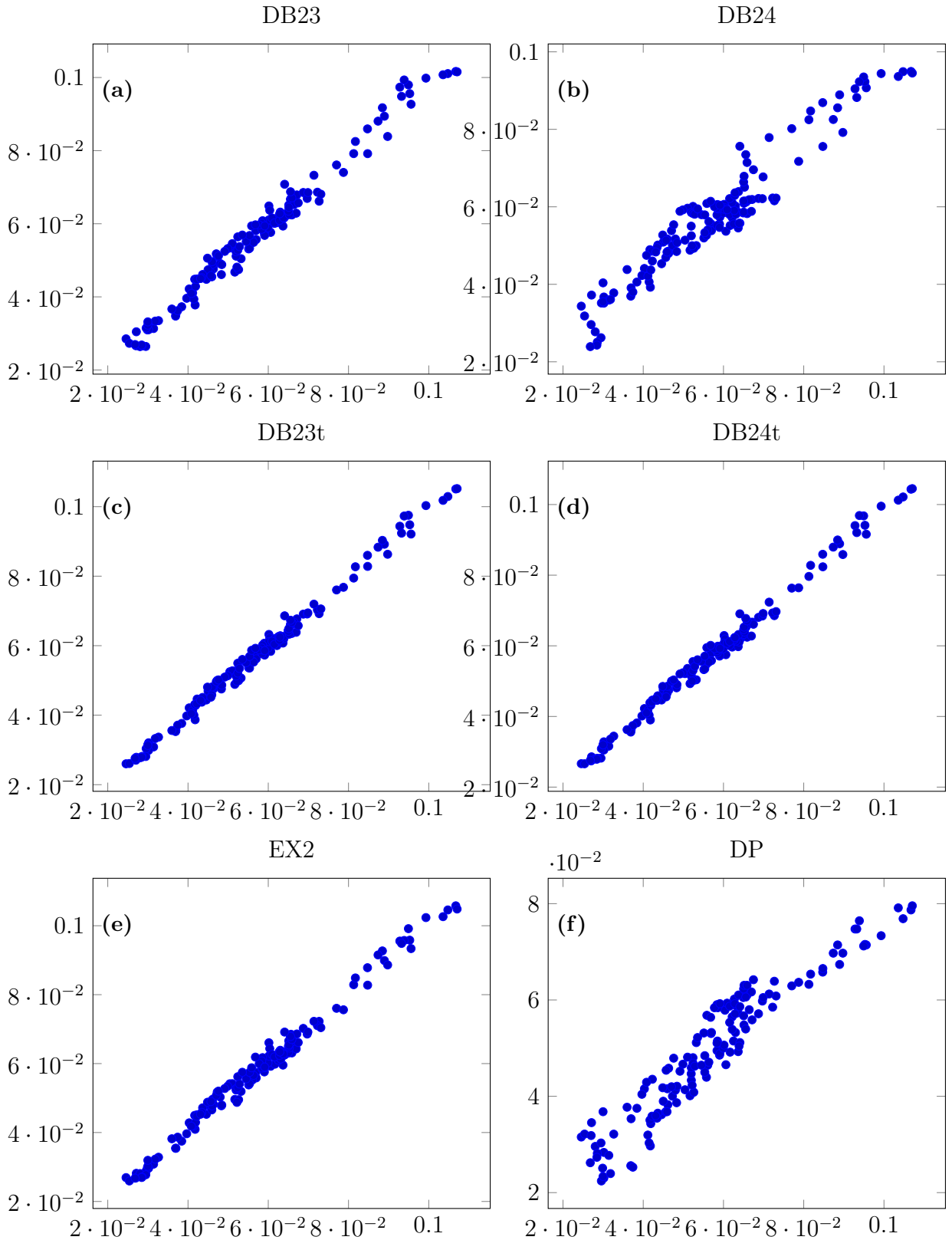
**Tabela 4 - Volatilidade relativa dos itens do IPCA - agosto/1999 a junho/2013**

Itens	Volatilidade relativa			
	08/1999 – 12/2002	01/2003 – 12/2007	01/2008 – 6/2013	08/1999 – 6/2013
Tubérculos, raízes e legumes	5,98	9,11	7,91	8,0
Hortaliças e verduras	4,49	5,09	4,80	4,8
Cereais, leguminosas e oleaginosas	3,30	3,26	3,64	3,4
Frutas	1,91	3,64	2,61	2,9
Açúcares e derivados	4,01	2,27	2,28	2,9
Aves e ovos	3,74	2,62	1,35	2,6
Combustíveis (veículos)	3,68	2,30	1,47	2,5
Fumo	2,08	1,40	3,08	2,4
Óleos e gorduras	3,25	1,89	1,85	2,3
Combustíveis (domésticos)	4,11	1,07	0,63	2,2
Leites e derivados	2,05	2,51	1,93	2,2
Pescados	2,27	2,22	1,95	2,1
Carnes	2,00	2,06	2,19	2,1

Obs: Os itens hachurados em cinza não são excluídos da medida de núcleo.

Fonte: Silva Filho e Figueiredo [54]

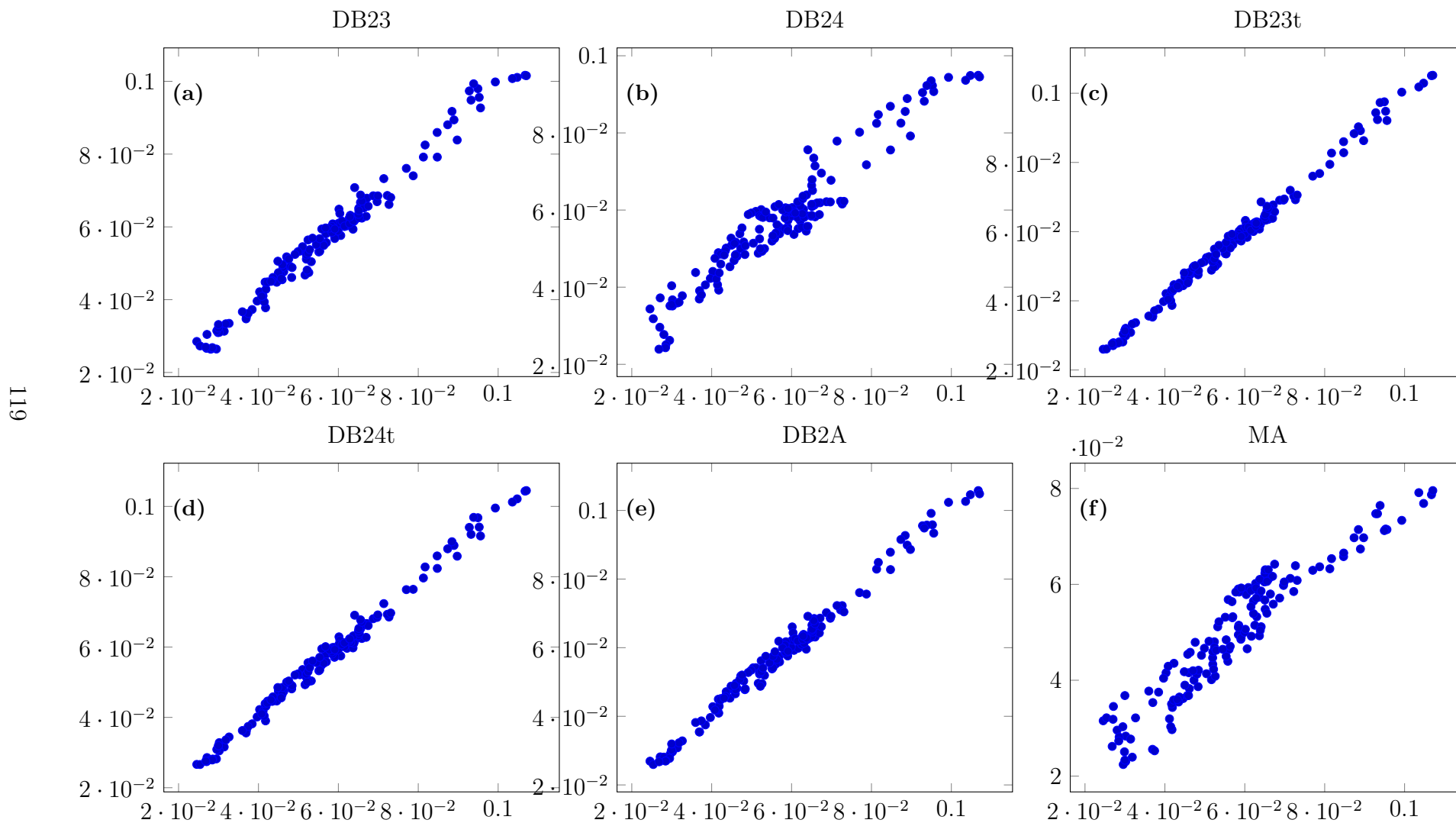
Figura 6.2: Scatter Plot Inflação/Núcleo



Fonte: elaborado pelo autor baseado em dados do BACEN [4] O eixo das abscissas representa o núcleo.

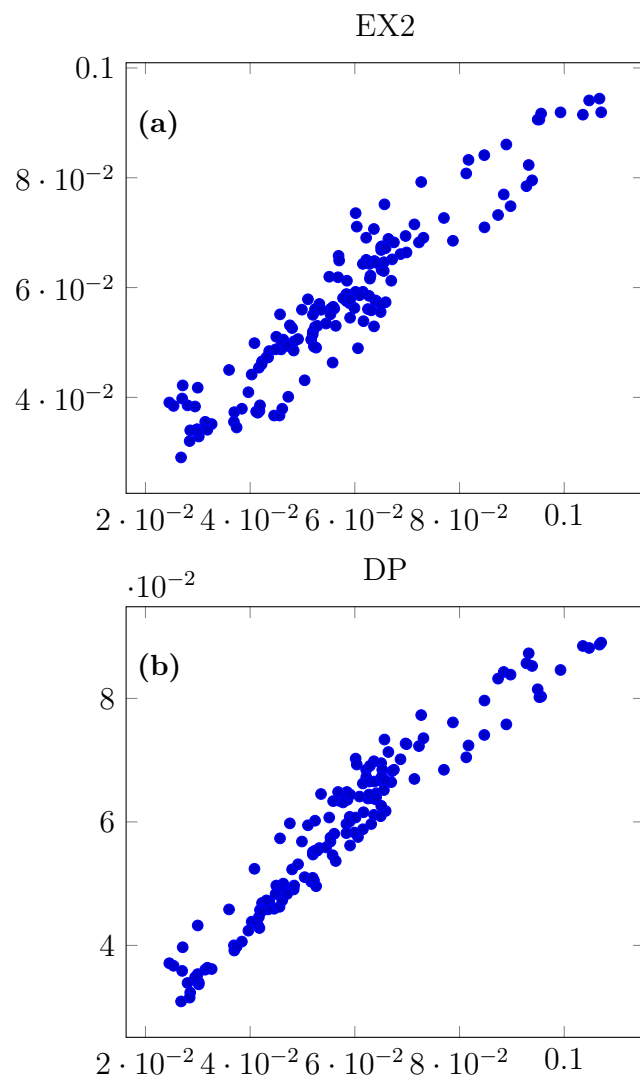


**Figura 6.3:** *Scatter plot* IPCA/núcleo com amostra restrita



Fonte: elaborado pelo autor baseado em dados do BACEN [4] O eixo das abscissas representa o núcleo.

**Figura 6.4:** Scatter plot IPCA/núcleo com amostra restrita



Fonte: elaborado pelo autor baseado em dados do BACEN [4] O eixo das abscissas representa o núcleo.