2
Modelos de Memória Longa

O fenômeno de memória longa foi primeiramente observado em séries hidrológicas e climatológicas no início dos anos de 1950, embora apenas na década de 80 os econometristas e estatísticos começaram a dar a devida atenção a este tipo de fenômeno.

Persistência ou *longa dependência* é o termo usado para descrever séries temporais, cuja função de autocorrelação exibe um comportamento que não é compatível com séries estacionárias, (integradas de ordem zero, $I(0)$) ou com séries não estacionárias (integradas de ordem 1, $I(1)$).

Neste tipo de processo, as autocorrelações amostrais entre lags distantes não podem ser consideradas desprezíveis, diferentemente de processos de memória curta, como processos $ARMA$, por exemplo, cujas observações separadas por um longo período de tempo são consideradas independentes. No domínio do tempo, a função de autocorrelação não é absolutamente somável apresentando o seguinte comportamento:

$$\sum_{k=-\infty}^{\infty} |\rho(k)| = \infty \tag{2.1}$$

Outra característica é que as autocorrelações $\rho(k)$ têm um decaimento lento e hiperbólico, equação 2.3, diferentemente do decaimento rápido e exponencial, equação 2.2, observado em séries tradicionais.

$$\rho_k \sim a^k \tag{2.2}$$

em que $0 < a < 1$

$$\rho(k) \sim e\,|k|^{2d-1}, |\mathrm{k}| \to \infty \tag{2.3}$$

No domínio da frequência, a presença de longa dependência pode ser observada através da função espectral que se torna ilimitada conforme a frequência tende a zero:

$$fx(\omega) \to \infty \qquad \text{quando} \qquad n \to 0 \tag{2.4}$$

SOUZA (1997) cita outras características deste tipo de série, dentre elas:

- Períodos longos em que as observações tendem a se manter em um nível baixo ou elevado;
- ao se observar a série de perto, isto é, ao se observar poucos anos, é possível perceber aparentes ciclos e/ou tendências, entretanto, ao se olhar a série como um todo (série composta de todas as observações possíveis), nenhum ciclo ou tendência parece persistir;
- A série aparenta estacionariedade;
- Variância tendendo a zero mais lento do que $n^{-\alpha}$, $0 < \alpha < 1.$;
- Correlações tendendo a zero aproximadamente proporcional a $k^{-\alpha}$, $0 < \alpha < 1.$;
- O gráfico log de periodograma *versus* log das frequências fica disperso em torno de uma linha reta decrescente.

Para se conseguir representar essas características, diversos modelos foram propostos. Um dos primeiros modelos a surgir foi o Ruído Gaussiano Fracionário (FGN), inicialmente proposto por MANDELBROT (1965), MANDELBROT & VAN NESS (1968) & MANDELBROT & WALLIS (1969). Embora este modelo não tenha alcançado um resultado satisfatório, a introdução do FGN no campo da hidrologia possibilitou o início do desenvolvimento teórico e prático dos modelos de memória longa.

Posteriormente, GRANGER & JOYEUX (1980) e HOSKING (1981) desenvolveram o modelo $ARFIMA(p,d,q)$. Este é uma generalização dos modelos $ARIMA$ para o caso em que d assume qualquer valor real, sendo um dos modelos de memória longa mais flexível e abrangente. Recentemente, estes modelos foram incorporados a livros de séries temporais e econométrios. Para mais detalhes, ver MORETTIN & TOLOI (2006), MILLS (1997), TSAY (2002), CHAN (2002), BROCKWELL & DAVIS (1991).

Os modelos utilizados neste trabalho pertencem à família dos modelos

ARFIMA e serão apresentados mais adiante. Para melhor entendimento dos mesmos, uma breve exposição sobre análise espectral será realizada na próxima seção.

2.1 Análise Espectral

Duas são as maneiras existentes na literatura para análise de séries temporais: a análise no domínio do tempo e a análise no domínio da frequência. No domínio do tempo, a inferência é realizada por meio da função de autocorrelação, que é uma boa ferramenta para avaliar a evolução do processo através do tempo. Analogamente, no domínio da frequência, existe a função densidade espectral, ou espectro, que tem como característica a representação de informações das influências de frequências sobre o processo. A análise espectral é o nome dado aos métodos utilizados para estimar a função densidade espectral.

Normalmente, a análise espectral é realizada quando o interesse principal é a busca por periodicidades ocultas na série. As suas maiores aplicações podem ser encontradas na engenharia elétrica, meteorologia, física e oceanografia, CHATFIELD (1996).

Suponha que $\{Z_t, t \in Z\}$ seja um processo estocástico estacionário com média zero e função de autocovariância satisfazendo uma condição de independência assintótica, no sentido de que valores muito distantes temporalmente sejam pouco dependentes, MORETTIN & TOLOI (2006). Desta maneira, a função densidade espectral é dada por:

$$f(\lambda) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\gamma_j(\tau)e^{-i\lambda\tau}, \ -\infty \leq \lambda \leq \infty, \tag{2.5}$$

Em que $e^{i\lambda} = \cos\lambda + isen\lambda$, $i = \sqrt{-1}$ e $\gamma(\tau)$ é a função de autocovariância. Esta função equivale à transformada de Fourier da função de autocovariância. Desta forma, entre as propriedades desta função, pode-se destacar que:

i. $f(\lambda)$ é contínua e possui valores reais não negativos, ou seja, $|f(\lambda)| = f(\lambda)$;

ii. $f(\lambda)$ é uma função periódica com período de 2π, $f(\lambda) = f(\lambda + 2\pi)$;

iii. $f(\lambda)$ é uma função par, ou seja, $f(\lambda) = f(-\lambda)$;

iv. A função γ_k pode ser obtida de $f(\lambda)$ através da transformada inversa de Fourier.

$$\gamma_k = \int_{-\infty}^{\infty} f(\lambda) e^{i\lambda k} d\lambda \tag{2.6}$$

De acordo com as propriedades (ii) e (iii), normalmente $f(\lambda)$ é representada no intervalo $(0, \pi)$, pois se conhecendo a função neste intervalo, todos os outros valores em toda a reta também serão conhecidos. A transformada inversa de Fourier é que permite recuperar γ_k a partir de $f(\lambda)$. A sequência de autocovariâncias $\{\gamma_k\}$ e a função $f(\lambda)$ formam um par de transformadas de Fourier. Desta forma, pode-se notar que a função de autocovariância e o espectro são ferramentas equivalentes, o que nos habilita a dizer que a análise no domínio de tempo e da frequência são análogos e a razão da escolha entre uma abordagem e outra está no objetivo da análise.

De posse de uma série temporal $\{X_t\}$, é de interesse obter uma estimativa de sua densidade espectral para que um modelo possa ser ajustado. Um estimador natural para a função densidade espectral é a função periodograma que é dada por:

$$I(\omega_j) = \frac{1}{2\pi}\left[\hat{\gamma}(0) + 2\sum_{s=1}^{n-1}\hat{\gamma}(s)\cos(s\omega)\right], \omega_j \in [-\pi, \pi] \tag{2.7}$$

em que $\hat{\gamma}(s)$ é a função de covariância amostral, $\omega_j = \frac{2\pi j}{N}$ são as frequências harmônicas, ou frequências de Fourier, com $j = 1, 2, ..., N/2$ em que N é o tamanho da série temporal. O periodograma também pode ser visto como uma função da transformada de Fourier discreta da função de autocorrelação amostral.

Embora o periodograma seja um estimador assintoticamente não viesado para a função densidade espectral, ele não é um estimador consistente. Em outras palavras, com o aumento do tamanho da série, a variância não tende à zero, CHATFIELD (1996), REISEN (2007).

Uma forma de solucionar tal problema, ou seja, obter um estimador não viesado e consistente para o espectro, é a inserção de janelas espectrais no periodograma. Este novo estimador é conhecido na literatura como periodograma suavizado. Em termos de transformada de Fourier, o periodograma suavizado nada mais é do que uma versão janelada da mesma por uma determinada função peso. Várias são as janelas sugeridas na literatura para a suavização do periodograma. Dentre elas pode-se citar a janela de Bartlett, a janela de Parzen, a janela de Tukey, a janela de Daniell, dentre outras, MORETTIN & TOLOI (2006).

Neste trabalho, será utilizada a janela de Parzen. O motivo de tal escolha é que a mesma não apresenta estimativas negativas para a densidade espectral, ver REISEN (1995). A janela de Parzen é definida da seguinte maneira:

$$\lambda(s) = \begin{cases} 1-6(s/m)^2+6(s/m)^3, |s|< m/2, \\ 2(1-(s/m))^3, m/2 \leq |s| \leq m, \\ 0, |s|> m \end{cases} \tag{2.8}$$

em que $m = n^{\beta}$ e $0 < \beta < 1$.

A escolha do parâmetro β é que vai determinar o nível de suavização que o periodograma irá sofrer. Quanto mais próximo de 1 for o valor de β, mais as estimativas do periodograma suavizado serão parecidas com as estimativas do periodograma, ou seja, o periodograma estará sofrendo uma pequena suavização. Por outro lado, quanto mais próximo de zero for o valor de β, mais forte será a suavização imposta ao periodograma. Desta forma, deve-se escolher β de modo que seja possível filtrar possíveis ruídos existentes no periodograma, sem que informações relevantes sejam perdidas devido a uma suavização excessiva.

O periodograma suavizado com a janela de Parzen é apresentado logo abaixo, em que $\lambda(s)$ é a janela espectral ou função de suavização:

$$f_s(\omega) = \frac{1}{2\pi} \sum_{s=-(n-1)}^{n-1} \lambda(s)\hat{\gamma}(s)\cos(s\omega) \tag{2.9}$$

Vale lembrar que o objetivo deste trabalho não é realizar uma análise de séries temporais no domínio da freqüência. Assim sendo, só foram apresentados os conceitos necessários para a elaboração e entendimento do mesmo. A utilização de tal ferramenta se dará no contexto da estimação dos parâmetros fracionários que serão introduzidos mais adiante. Mais detalhes sobre séries temporais no domínio da frequências podem ser encontrados em HAMILTON (1994) e CHATFIELD (1996).

2.2
O Modelo *ARFIMA*

Os modelos *ARFIMA* (*Autoregressive Fractional Integrated Moving Average*) são uma generalização dos modelos $ARIMA\,(p,d,q)$, em que o número de diferenciações d assume valores não inteiros. Desta forma, d passa a ser um parâmetro desconhecido e, portanto, deve ser estimado. Este fato é que permite a incorporação dos efeitos de memória longa aos modelos $ARMA\,(p,q)$ tradicionais.

A principal vantagem destes modelos é a flexibilidade de modelar estruturas de autocorrelação da série, tanto nos *lags* altos (memória longa), como nos *lags* baixos (memória curta). Para isso, deve-se estimar o parâmetro d de forma a captar a estrutura de correlação nas ordens altas, enquanto que os parâmetros ϕ e θ devem ser ajustados com o intuito captar os efeitos das ordens baixas da série.

Por outro lado, a estimação dos parâmetros, especialmente o fracionário, é a principal desvantagem do método, uma vez que existem várias abordagens e diversos estimadores, tanto no domínio do tempo, quanto no domínio da frequência. É possível encontrar diversos estudos comparando os estimadores disponíveis na literatura, REISEN, ABRAHAM & TOSCANO (2000), LOPES, OLBERMAN & REISEN (2003), TREVISAN, SOUZA & SOUZA (2000). No entanto, ainda não existe um consenso na literatura sobre qual é a melhor abordagem ou o melhor estimador.

Formalmente, tem-se $\{X_t\}$ dito um processo $ARFIMA(p,d,q)$ se puder ser escrito da seguinte forma:

$$\phi(B)(1-B)^d(X_t-\mu)=\theta(B)a_t \tag{2.10}$$

em que $d \in R$, $\phi(B)=1-\phi_1 B-\phi_2 B^S-\ldots-\phi_p B^p$ é o polinômio autorregressivo de ordem p, $\theta(B)=1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q$ é o polinômio de média móvel de ordem q e a_t é um ruído branco.

O operador de retardo, o termo $(1-B)^d$, é dado através da expansão binomial da seguinte forma:

$$(1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k = 1 - dB - \frac{d}{2!}(1-d)B^2 - \ldots \tag{2.11}$$

HOSKING (1981) prova que X_t é inversível se $d > -1/2$ e se as raízes de $\theta(B)=0$ estiverem fora do círculo unitário. X_t é estacionário de $d < 1/2$ e as raízes de $\phi(B)=0$ estiverem fora do círculo unitário. Nestas condições, a função densidade espectral deste processo é dada por:

$$fx(w) = f_u(w)(2sen(w/2))^{-2d}, w \in [-\pi, \pi] \tag{2.12}$$

em que $f_u(w)$ é a função densidade espectral de um processo $ARMA\ (p,q)$.

Vale lembrar que a característica de memória longa é observada quando $0 < d < 1/2$. Quando $-1/2 < d < 0$, é utilizado na literatura o termo memória intermediária.

Quando $d = 0$, o modelo $ARFIMA$ é reduzido a um modelo $ARMA$ padrão. Se $p = q = 0$, ou seja, $ARFIMA\ (0,d,0)$, este recebe o nome de modelo de diferenciação fracionária, e matematicamente, tem-se:

$$\nabla^d X_t = a_t \tag{2.13}$$

Como dito anteriormente, a fase crítica na construção de um modelo $ARFIMA$ é a estimação dos seus parâmetros.

As principais formas existentes na literatura para proceder esta estimação são o método paramétrico e o método semiparamétrico. No método paramétrico, todos os parâmetros do modelo são estimados ao mesmo tempo, ou seja, tanto os parâmetros ϕ e θ, quanto o parâmetro d são estimados simultaneamente. Neste tipo de estimação pode-se citar os estimadores propostos por FOX & TAQQU (1986), DAHLHAUS (1989), SOWELL (1992).

Com relação ao método semiparamétrico, os parâmetros do modelo são estimados em momentos diferentes. Inicialmente estima-se o parâmetro fracionário d e posteriormente estimam-se os parâmetros ϕ e θ. Dentre os métodos semiparamétricos, pode-se citar o proposto por GEWEKE & PORTER-HUDAK (1983), REISEN (1994), ROBINSON (1995).

O procedimento semiparamétrico deve ser realizado da seguinte forma:

- Estima-se o parâmetro fracionário *d*;
- Verifica-se a significância estatística do mesmo, ou seja, se o parâmetro estimado é diferente de zero;
- Caso o parâmetro seja diferente de zero, deve-se realizar uma diferenciação fracionária na série original. Supondo que a série seja X_t e que esta siga um $ARFIMA(p,d,q)$, deve-se fazer $\widehat{U}_t = (1-B)^d (X_t)$;
- Em posse desta nova série $\widehat{U}_t$, livre do componente de memória longa, deve-se estimar os parâmetros θ e ϕ do processo $ARMA$;
- Após a estimação de todos os parâmetros, é necessário averiguar os resíduos do modelo.

A descrição deste procedimento também pode ser encontrada em REISEN & LOPES (1999), MOLINARES, REISEN & CRIBARI-NETO (2007).

Neste trabalho utilizou-se a construção semiparamétrica, devido à facilidade de implementação e o baixo custo computacional.

2.2.1 Método de Regressão do Parâmetro Fracionário *d*

Seja $\{X_t\}$ um processo $ARIMA\,(p,d,q)$ com $d \in (-0.5, 0.5)$ concebido na equação (2.10) e com função densidade espectral de X_t dada pela expressão (2.12). Aplicando o logaritmo em (2.12), obtém-se:

$$\ln f(\omega) = \ln f_u(\omega) - d\ln(2sen(\omega/2))^2 \qquad (2.14)$$

Esta equação também pode ser escrita de outra forma:

$$\ln f(\omega) = \ln f_u(0) - d\ln[2sen(\omega/2)]^2 + \ln[f_u(\omega)/f_u(0)] \tag{2.15}$$

A expressão (2.15) será empregada nas duas próximas subseções para estimar d no modelo $ARFIMA(p,d,q)$.

2.2.1.1
Método de Regressão Utilizando a Função Periodograma

Este estimador foi proposto por GEWEKE & PORTER-HUDAK (1983) e baseia-se em uma equação que exibe a relação entre a função densidade espectral de um processo $ARFIMA(p,d,q)$ e de um processo $ARMA\,(p,q)$.

Retornando à equação (2.15), substituindo ω pelas frequências de Fourier $\omega_j = \frac{2\pi j}{n}$, e adicionando $\ln I(\omega_j)$ em que $I(\cdot)$ é uma função periodograma definida em (2.7), obtém-se:

$$\ln I(\omega_j) = \ln f_u(0) + \ln\left[\frac{f_u(\omega_j)}{f_u(0)}\right] + \ln\left[\frac{I(\omega_j)}{fx(\omega_j)}\right] - d\ln\left[2sen\left(\frac{\omega_j}{2}\right)\right]^2 \tag{2.16}$$

Adotando o limite superior de j igual a $g(n)$, que deve ser escolhido atendendo à condição $\frac{g(n)}{n} \to 0$ quando $n \to \infty$, o termo $\ln\left[\frac{f_u(\omega_j)}{f_u(0)}\right]$ da expressão (2.16) pode ser considerado desprezível comparado aos outros termos. GEWEKE & PORTER-HUDAK (1983) sugerem que $g(n) = n^{\alpha}$, sendo n o número de observações da série temporal e α uma constante entre 0 e 1.

Assim sendo, a equação (2.16) pode ser aproximada como

$$\ln I(\omega_j) \cong \ln f_u(0) - d\ln\left[2sen\left(\frac{\omega_j}{2}\right)\right]^2 + \ln\left[\frac{I(\omega_j)}{fx(\omega_j)}\right]. \tag{2.17}$$

Esta expressão é similar a uma regressão linear simples da forma

$$Y_j = a + bX_j + e_j \qquad j = 1,2,\ldots,g(n) \tag{2.18}$$

em que,

$$Y_j = \ln I(\omega_j) \qquad (2.19)$$

$$a = \ln f_u(0) - c \qquad (2.20)$$

$$b = -d \qquad (2.21)$$

$$X_j = \ln\left[2sen\left(\frac{\omega_j}{2}\right)\right]^2 \qquad (2.22)$$

$$e_j = \left[\ln\frac{I(\omega_j)}{fx(\omega_j)}\right] + c \qquad (2.23)$$

$$c = E\left[-\ln'\frac{I(\omega_j)}{fx(\omega_j)}\right] \qquad (2.24)$$

REISEN (2007) afirma que as variáveis $[\ln I(\omega_j)/fx(\omega_j)]$, $j = 1,2,...,g(n)$ são aproximadamente independentes de distribuição Gumbel com média igual a -0,577216 e variância igual a $\pi^2/6$. Assim sendo, os erros e_j também são aproximadamente independentes de distribuição Gumbel, com média igual a zero e a variância $\pi^2/6$.

Isto posto, utilizando mínimos quadrados ordinários na regressão de $Y_1, Y_2,..., Y_{g(n)}$ em $X_1, X_2,..., X_{g(n)}$, obtém-se o estimador para $\hat{d}$.

$$\hat{d} = -\frac{\sum_{j=1}^{g(n)}(X_j - \overline{X})Y_j}{\sum_{j=1}^{g(n)}(X_j - \overline{X})^2} \qquad (2.25)$$

Como pode ser observado na equação (2.25), o valor estimado para o parâmetro de longa dependência é o valor do coeficiente da regressão multiplicado por menos 1.

$$\hat{d} = -\widehat{b} \qquad (2.26)$$

Segundo GEWEKE & PORTER-HUDAK (1983), o estimador $\hat{d}$ do parâmetro fracionário tem distribuição assintoticamente normal quando

$d \in (-0.5, 0)$ com as seguintes propriedades:

$$E(\hat{d}) = d \tag{2.27}$$

$$Var(\hat{d}) = \frac{\pi^2}{6\sum_{j=1}^{g(n)}(X_j - \overline{X})^2} \tag{2.28}$$

2.2.1.2
Método de Regressão Utilizando a Função Periodograma Suavizado

Este estimador foi proposto por REISEN (1994) e consiste em utilizar a função periodograma suavizado (expressão 2.9) no lugar da função periodograma (expressão 2.7) presente na equação 2.16.

Como dito no item 2.1, o motivo da permuta é que a função periodograma não é um estimador consistente da função densidade espectral.

Retornando a equação 2.16 e 2.17, e trocando a função periodograma pela função periodograma suavizado, obtém-se:

$$\ln f_s(\omega_j) = \ln f_u(0) + \ln\left[\frac{f_u(\omega_j)}{f_u(0)}\right] + \ln\left[\frac{f_s(\omega_j)}{fx(\omega_j)}\right] - d\ln\left[2sen\left(\frac{\omega_j}{2}\right)\right]^2 \tag{2.29}$$

Limitando $j = 1,2,..., g(n)$ e escolhendo $g(n) = n^\alpha$, em que $0 < \alpha < 1$, a equação 2.29 pode ser redefinida como:

$$\ln f_s(\omega_j) \cong \ln f_u(0) - d\ln\left[2sen\left(\frac{\omega_j}{2}\right)\right]^2 + \ln\left[\frac{f_s(\omega_j)}{fx\left(\omega_j\right)}\right]. \tag{2.30}$$

A equação (2.30) pode ser aproximada por uma regressão linear simples.

$$Y_j = a + bX_j + \varepsilon_j \qquad j = 1,2,..., g(n) \tag{2.31}$$

em que,

$$Y_j = \ln f_s(\omega_j) \tag{2.32}$$

$$a = \ln f_u(0) \tag{2.33}$$

$$b = -d \tag{2.34}$$

$$X_j = \ln\left[2sen\left(\frac{\omega_j}{2}\right)\right]^2 \tag{2.35}$$

$$\varepsilon_j = \left[\frac{f_s(\omega_j)}{fx(\omega_j)}\right] \tag{2.36}$$

ZAMPROGNO (2004) apud REISEN (1995) diz que os erros ε_j possuem distribuição normal com média zero e variância igual a

$$Var\left[\ln\left(\frac{f_s(\omega)}{f(\omega)}\right)\right] \sim \begin{cases} 0{,}53928\left(\frac{m}{n}\right), & \omega \neq 0, \pi \\ 1{,}07856\left(\frac{m}{n}\right), & \omega = 0, \pi \end{cases} \tag{2.37}$$

Com isto, o estimador de $\hat{d}_{ps}$ proposto por REISEN (1994) é dado por

$$\hat{d}_{ps} = -\frac{\sum_{j=1}^{g(n)}(X_j - \overline{X})Y_j}{\sum_{j=1}^{g(n)}(X_j - \overline{X})^2} \tag{2.38}$$

e, como no estimador utilizando a função periodograma, o valor estimado para $\hat{d}$ consiste no valor do coeficiente da regressão multiplicado por menos 1.

$$\hat{d}_{ps} = -\hat{b} \tag{2.39}$$

Este estimador apresenta as seguintes propriedades assintóticas:

$$E\left(\hat{d}_{ps}\right) = d \tag{2.40}$$

$$Var(\hat{d}_{ps}) = 0{,}53928\frac{m}{n\sum_{j=1}^{g(n)}(X_j - \overline{X})^2}, \quad \omega \neq 0, \pi \tag{2.41}$$

Embora o *ARFIMA* não tenha sido empregado na dissertação, é de extrema importância o entendimento do mesmo, pois é através deste e da estimação descrita que será possível a extensão matemática do modelo a ser implementado.

2.3
O Modelo SARFIMA

Recentemente, a metodologia utilizada em séries temporais com longa dependência foi ampliada para séries temporais com sazonalidade. Uma série temporal exibe um fenômeno sazonal quando este se repete regularmente depois de período de tempo não superior a um ano. O menor período de tempo desta repetição do fenômeno é denominado período sazonal.

O estudo da componente sazonal considerando o fenômeno de longa dependência é bem recente. Um dos trabalhos pioneiros foi o de PORTER-HUDAK (1990), que examinou um conjunto de séries de agregados monetários. Para isso, a autora ampliou o modelo *ARFIMA* para o caso sazonal.

PEIRES & SINGH (1996) definem o processo $SARFIMA\{X_t\}$ como:

$$\Phi(B^s)\phi(B)(1-B)^d(1-B^s)^D X_t = \theta(B)\Theta(B^S)a_t \tag{2.42}$$

Os polinômios autorregressivos simples e sazonais, médias móveis simples e sazonais, são respectivamente $\phi(B) = 1 - \phi_1 B - \phi_2 B^S - \ldots - \phi_p B^p$, $\Phi(B^s) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \ldots - \Phi_P B^{PS}$, $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$ e $\Theta(B^s) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \ldots - \Theta_Q B^{QS}$. As ordens dos polinômios p, q, P, Q assumem valores inteiros e positivos enquanto que os parâmetros fracionários *d* e *D* assumem valores reais.

O operador de diferenciação fracionária simples, $(1-B)^d$, é definido como na equação 2.11. O operador de diferenciação fracionária sazonal, $(1-B^S)^D$, é da forma:

$$(1-B^S)^D = 1 - DB^S - \frac{D(1-D)}{2!}B^{2S} - \frac{D(1-D)(2-D)}{3!}B^{3S} - \ldots \qquad (2.43)$$

O modelo é estacionário e invertível se e somente se, $|d+D|<1/2$, $|D|<1/2$, $|d|<1/2$ e $\phi(B),\theta(B),\Phi(B)$ e $\Theta(B)$ têm suas raízes fora do círculo unitário. O processo possui características de memória longa quando $|d+D|>0$.
A função densidade espectral do modelo da equação 2.42 é dada por:

$$fx(\omega) = fu(\omega)\left[2sen(\omega s/2)\right]^{-2D}\left[2sen(\omega/2)\right]^{-2d} \qquad (2.44)$$

para $-\pi \le w \le \pi$, em que $fu(\cdot)$ é a densidade espectral do processo $SARMA$.

Definido o modelo que usado neste trabalho, segue a apresentação do método de estimação.

2.3.1 Estimadores de *d* & *D*

Desde o trabalho de GEWEKE & PORTER-HUDAK (1983), muitos estimadores foram propostos na literatura para o parâmetro fracionário d de um modelo $ARFIMA$. Com a extensão dos modelos $ARFIMA$ para os modelos $ARFIMA$ sazonais, ou $SARFIMA$, muitos desses estimadores também foram adaptados para o caso sazonal.

O método de estimação utilizado nesta dissertação é similiar ao método proposto por REISEN, RODRIGUES & PALMA (2006a, b), que é baseado no método proposto por GEWEKE & PORTER-HUDAK (1983). No entanto, assim como REISEN (1994) sugeriu a substituição da função periodograma pela função periodograma suavizada no método GPH não sazonal, também será realizada esta adaptação para o caso sazonal. Isto equivale a dizer que estamos utilizando uma expansão do método proposto por REISEN (1994) para o caso em que a sazonalidade está presente nos dados.

Assim como no caso da estimação do parâmetro de memória longa no $ARFIMA(p,d,q)$, este método de estimação também consiste em uma regressão usando a função densidade espectral (equação 2.44) para montar uma regressão entre o periodograma suavizado e as frequências harmônicas.

Para isto, deve-se logaritmizar a equação 2.44, obtendo assim:

$$\ln fx(\omega) = \ln fu(\omega) - D\ln\left[2sen(\omega s/2)\right]^2 - d\ln\left[2sen(\omega/2)\right]^2 \qquad (2.45)$$

A função periodograma ou periodograma suavizada deve ser utilizada como estimativa para $fx(\cdot)$. Caso seja utilizada função periodograma, tem-se o estimador proposto por REISEN, RODRIGUES & PALMA (2006a, b). Por outro lado, ao se utilizar a função periodograma suavizada obtem-se o método proposto por REISEN (1994) ampliado para o caso sazonal.

A equação 2.45 é similar a uma regressão linear múltipla da forma:

$$Y_j = a + b_1X_{1j} + b_2X_{2j} + \varepsilon_j \qquad (2.46)$$

em que,

$$Y_j = \ln I(\omega_j) \qquad (2.47)$$

$$X_{1j} = \ln\left[2sen(\omega_j s)\right]^2 \qquad (2.48)$$

$$X_{2j} = \ln\left[2sen(\omega_j/2)\right]^2 \qquad (2.49)$$

$$b_1 = -D \qquad (2.50)$$

$$b_2 = -d \qquad (2.51)$$

$$\varepsilon_j = \ln\frac{I(\omega_j)}{f(\omega_j)} - E\left[\ln\frac{I(\omega_j)}{f(\omega_j)}\right] \qquad (2.52)$$

Desta maneira, $\hat{d}$ e $\hat{D}$ devem ser estimados através de mínimos quadrados ordinários. Como pode ser observada em 2.50 e 2.51, a estimativa dos parâmetros fracionários é o valor encontrado para os coeficientes da regressão (b_1 e b_2) multiplicado por menos um.

Cabe salientar também que o número de observações na regressão é determinado por $g(n) = n^{\alpha}$, em que α é uma constante entre zero e um e n é o tamanho da série temporal. Além disto, $g(n)$ deve satisfazer a condição

$\left(\frac{g(n)}{n}\right)\log g(n)+\frac{1}{g(n)}\to 0$ quando $g(n)\to\infty$ e $n\to\infty$. Em termos práticos, pode-se dizer que a regressão é realizada com $j=1,2,\ldots,g(n)$.

Algumas propriedades assintóticas do estimador que utiliza a função periodograma como estimador de $fx(\cdot)$ podem ser encontradas em GRIPA (2008).

O estimador usado nesta dissertação foi a função periodograma suavizado e, desta forma, os resultados expostos em GRIPA (2008) não podem ser utilizados. Uma das maneiras de solucionar este fato é a aplicação do *bootstrap* inserido no próximo capítulo, para a obtenção de uma distribuição empírica dos parâmetros fracionários.