

3 Metodologia

Neste capítulo serão apresentados os procedimentos de estimação utilizados no levantamento das estimativas dos parâmetros do modelo log-periódico para a série de preços dos índices financeiros. Além disso, será feita uma investigação econométrica do modelo log-periódico, descrevendo os potenciais problemas inerentes ao mesmo.

3.1. Premissas Gerais

Inicialmente, serão listadas as premissas gerais que envolvem as restrições sobre os parâmetros. Estas, por serem de origem conceitual, se sobrepõe aos procedimentos de estimação, ou seja, atuam como premissas globais.

Uma série de estudos empíricos, [1], [2], [3], [7], indicaram um conjunto de premissas sobre os parâmetros a serem estimados do modelo log-periódico de evolução do preço do ativo financeiro dado por:

$$p(t) = A + B(t_c - t)^\beta + C(t_c - t)^\beta \cos[w \ln(t_c - t) + \phi] \quad (42)$$

- $A \sim p(t_c) \rightarrow$ o valor do parâmetro A é aproximadamente da ordem do valor máximo do preço no instante imediatamente anterior ao crash [6]. Analisando o comportamento de todas as séries utilizadas nesta dissertação, definiu-se que o parâmetro A deve pertencer ao intervalo [300,30000];
- $B < 0 \rightarrow$ o parâmetro B deve ser negativo quando os dados indicarem uma bolha terminando em um crash [3];
- $C \neq 0 \rightarrow$ o parâmetro C deve ser não nulo para garantir a consistência da eq.(42) com o modelo da imitação racional em um sistema hierárquico e, conseqüentemente com as propriedades de invariância por escala do modelo;

Além disso, $|C| < |B|$ para que a componente log-periódica em (42) represente uma correção ao crescimento em lei de potência pura do preço.

- O crash financeiro é um evento que pode durar de alguns dias a algumas semanas. Desta forma, o t_c previsto pelo modelo deve estar contido entre $t_{m\acute{a}x}$ e $t_{m\acute{i}n}$, onde $t_{m\acute{a}x}$ é o tempo no qual o índice financeiro atinge seu maior valor antes do crash e $t_{m\acute{i}n}$ é o tempo no qual o índice financeiro atinge seu valor mais baixo após o crash. Como conseqüência deste fato, o intervalo de busca para o parâmetro t_c foi definido entre 1 dia e 0.2 anos após a última observação dos dados [3];
- Estudos empíricos envolvendo uma ampla variedade de sistemas, como por exemplo em processos de crescimento, rupturas, terremotos, crashes financeiros, sugerem que o fator de invariância por escala λ característico da rede hierárquica pertence ao intervalo (2,3) [7]. Conseqüentemente pela eq.(41), define-se $5 < w < 15$, o que corresponde a $1.5 < \lambda < 3.5$; Deve-se notar que, para valores de w muito pequenos, não se observaria uma oscilação completa em todo o intervalo, significando pouca representatividade desta modelagem. Por outro lado, para valores de w muito grandes, seriam observadas muitas oscilações, conseqüentemente fitando-se o ruído [1];
- $0 < \beta < 1$ por razões econômicas, apresentadas na seção 2.3.1; Além disso, β não deve estar muito próximo de 0 ou de 1: valores de β muito pequenos, representam uma bolha plana, com aceleração repentina próxima ao crash; valores de β próximos de 1, representam uma bolha não acelerada [1].
- Nenhum estudo anterior mencionou a necessidade de alguma restrição sobre o parâmetro ϕ . No entanto, neste trabalho ϕ será restrito entre 0 e 2π , que representa um ciclo completo da função periódica cosseno.

Todos os parâmetros $(A, B, C, t_c, \beta, w, \phi)$ restritos entre a e b listados acima foram re-parametrizados de acordo com a transformação monotônica:

$$\Psi = b \frac{\exp(\varphi)}{1 + \exp(\varphi)} + a \left(1 - \frac{\exp(\varphi)}{1 + \exp(\varphi)} \right)$$

onde $\varphi \in (-\infty, +\infty)$. Esta transformação se fez necessária para que a estimação ocorresse sobre um espaço de solução irrestrito, o que diminui a complexidade da otimização.

A variável t (tempo) foi convertida em unidades de anos: sendo um ano financeiro equivalente a 252 dias de pregão, tem-se que 1 dia = 0.003968 de ano. Da mesma forma, 0.01 anos = 2.52 dias e 0.1 anos = 25.2 dias (que

equivale a aproximadamente 5 semanas). Esta conversão foi utilizada nos cálculos do parâmetro t_c .

Para estimar o modelo, foram utilizados dois algoritmos de otimização: um determinístico, BFGS (Broyden – Fletcher – Goldfarb – Shanno) [8] e um heurístico, *Generalized Simulated Annealing* (GSA) [9].

Em ambos os casos, a solução ótima foi obtida através da minimização de uma função custo pré-definida pelo usuário, considerando-se um espaço de solução irrestrito.

A implementação do BFGS utilizou rotinas do Matlab, já a implementação do GSA utilizou rotinas em linguagem de programação C.

O procedimento utilizado na estimação de nosso modelo será ilustrado através da análise da série de preços do índice financeiro do mercado americano S&P500, que corresponde a um período que inclui a ocorrência do *crash* de 1987.

A figura abaixo mostra a série do índice de preços do S&P500 no período 1980 – 1990, na qual é observada queda significativa do preço do ativo a partir de agosto de 1987, culminando com queda acentuada em outubro de 1987.

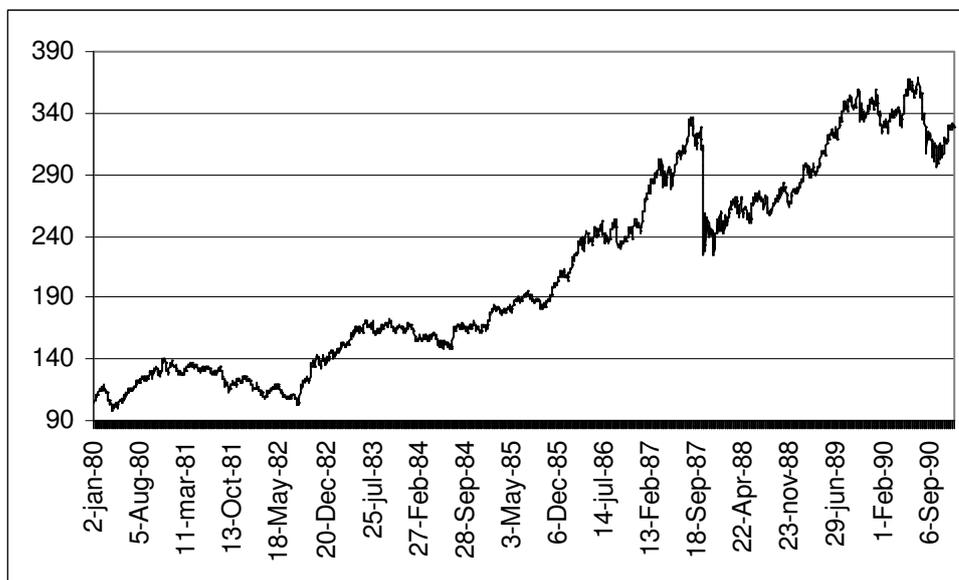


Figura 16 – Série de preços do S&P500 de Jan-1980 a Dez-1990.

Para procedermos com a estimação do modelo é necessário identificar a bolha especulativa pré-*crash* caracterizada por uma aceleração prolongada dos preços. O final da bolha é determinado de forma objetiva pela data de $t_{máx}$ onde o maior valor do índice é obtido antes do *crash*. No entanto, o início da bolha não tem determinação tão clara, sendo identificado como a data do menor valor do

índice associado à mudança de tendência dos preços. Desta forma, o período utilizado em nossa análise foi de 01/07/85 a 25/08/87.

3.2.

Investigação Econométrica do Modelo Log-periódico: Estimação do Modelo em Nível

Nesta primeira etapa, será estimado o modelo log-periódico no nível $p(t)$ para a série de preços do S&P500. Serão apresentados os resultados das estimativas obtidas, além de uma análise crítica do modelo log-periódico segundo critérios econométricos.

Dentre os potenciais problemas econométricos apresentados na literatura sobre os modelos de regressão estão: a não-identificação dos parâmetros; a estimação em séries não estacionárias, como é o caso da a série de preços dos índices financeiros; a presença de autocorrelação e heterocedasticidade nos resíduos do modelo e a inferência possível em um modelo com tendência determinística não-linear [10], [11], [12].

3.2.1.

Identificação

Um modelo de regressão não linear é dito identificado se, para um conjunto específico de dados, obtém-se um único vetor de parâmetros estimados $\hat{\Phi}$ que minimize o erro quadrático médio. Esta definição também é denominada de identificação global. Assim, se o modelo é não identificado pelo conjunto de dados utilizado na estimação, significa que existe pelo menos mais de um valor ótimo para Φ [10].

Re-escrevendo o modelo log-periódico, eq.(42), da seguinte forma:

$$p(t) = A + B\eta_1(t, \varphi_1) + C\eta_2(t, \varphi_2) + \varepsilon_t \quad (43)$$

sendo $\eta_1(t, \varphi_1)$ e $\eta_2(t, \varphi_2)$ as representações funcionais das componentes que multiplicam os parâmetros B e C, respectivamente, na expressão original do modelo, e onde $\varphi_1 \equiv [t_c, \beta]$ e $\varphi_2 \equiv [t_c, \beta, w, \phi]$.

Note que o modelo log-periódico será não identificado para $\hat{C} = 0$, pois, neste caso, qualquer valor de φ_2 poderá ser solução e conseqüentemente haverá infinitos valores ótimos para o vetor de parâmetros φ_2 . Da mesma

forma, o modelo será não identificado para $\hat{B} = 0$; e também será não identificado para $\hat{\varphi}_2 = 0$, pois, neste caso, a expressão do modelo se resumirá à componente constante. Por último, o modelo será não identificado para $\hat{\phi}$, devido à periodicidade da própria função cosseno, onde todos os múltiplos inteiros $\hat{\phi} + 2n\pi$ possuem a mesma imagem.

Logo, a princípio, a expressão do modelo log-periódico, eq.(43), apresenta problemas de identificação.

Entretanto, de acordo com as premissas gerais estabelecidas na seção 3.1, verifica-se que nenhum dos casos de não identificação, apresentados acima, são soluções possíveis para os valores verdadeiros dos parâmetros, já que o espaço de solução viável não contém $\varphi_2 = 0$ ou $C = 0$ ou $B = 0$ ou múltiplos de $\hat{\phi}$.⁹

Assim, o modelo log-periódico quando aplicado a dados de índices financeiros é, por construção, um modelo globalmente identificado, pois o espaço de solução viável não contempla os casos-problema.

3.2.2. Estimativas do Modelo Log-periódico

Nesta seção será apresentado o procedimento de levantamento das estimativas do vetor de parâmetros ótimo $\hat{\Phi}$ do modelo log-periódico, onde $\Phi = [A, B, C, t_c, \beta, w, \phi]$. Serão consideradas as premissas gerais, descritas na seção 3.1.

A função custo foi definida como a soma dos quadrados dos resíduos, $S(\Phi) = \sum_{t=1}^n \hat{u}_t^2(\Phi)$, onde $\hat{u}_t(\Phi) = p_t - \hat{p}_t(\Phi)$, sendo p_t os valores observados e $\hat{p}_t(\Phi)$ a série de preços prevista pelo modelo log-periódico segundo a eq.(42).

⁹ A restrição imposta sobre $\hat{\phi}$ na seção 3.1 foi justamente para eliminar a não identificação do modelo. Esta restrição pode ser imposta, pois não está em desacordo com a teoria e não altera, em nenhuma situação, os valores possíveis da função.

As estimativas para $\hat{\Phi}$, obtidas através da minimização de $S(\Phi)$, são os estimadores de mínimos quadrados ordinários (MQO) para Φ .

Devido ao comportamento de $S(\Phi)$ altamente degenerado ou quase-degenerado, a obtenção de uma solução ótima tornou-se não-trivial. Desta forma, a escolha da técnica de otimização é de suma importância e, em particular, deve ser eficaz em evitar mínimos locais.

Com este objetivo, inicialmente, será utilizado como otimizador uma rotina baseada na técnica de GSA [9]. Esta técnica heurística permite visitar o espaço do vetor de parâmetros Φ com passos estocásticos de alcance arbitrário, permitindo assim uma busca através de regiões associadas a diferentes mínimos locais. O resultado fornecido por este otimizador é então utilizado como condição inicial para a aplicação do algoritmo BFGS.

Os resultados obtidos para $\hat{\Phi}$ estão mostrados na tabela a seguir:

Crash SP&500 - 1987	
	Log-per
A	399.43
B	-153.06
C	-12.09
β	0.35
w	7.28
φ	1.19
t_c	2.239

Tabela 1 - Parâmetros estimados do modelo log-periódico para o S&P500 – 1987.

Deve-se notar que os parâmetros ótimos obtidos satisfazem às restrições apresentadas na seção anterior. Em particular, o tempo mais provável para ocorrência do *crash*, t_c , está contido entre t_{\max} e t_{\min} , conforme mostrado na tabela a seguir:

Crash SP500 - 1987						
	t_c	Data t_c	t_{\max}	Data t_{\max}	t_{\min}	Data t_{\min}
Log-per	2.239	23-Sep-87	2.159	25-Aug-87	2.310	19-Oct-87

Tabela 2 – Valor de t_c previsto pelo modelo e valores empíricos de t_{\max} e t_{\min} .

A figura abaixo apresenta a curva log-periódica estimada superposta à série empírica do S&P500.

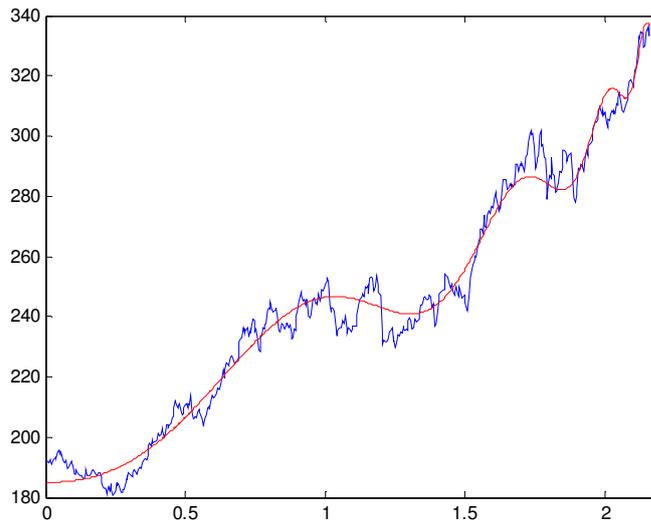


Figura 17 - Em azul série do S&P500 observada, em vermelho série estimada pelo modelo log-periódico.

Foi também obtida a série de resíduos \hat{u}_t , onde $\hat{u}_t = p_t - \hat{p}_t(\hat{\Phi})$, sendo p_t os valores observados e $\hat{p}_t(\hat{\Phi})$ as estimativas do modelo log-periódico segundo a eq.(42) e os parâmetros ótimos da tabela 1. A série de resíduos está mostrada no gráfico abaixo.

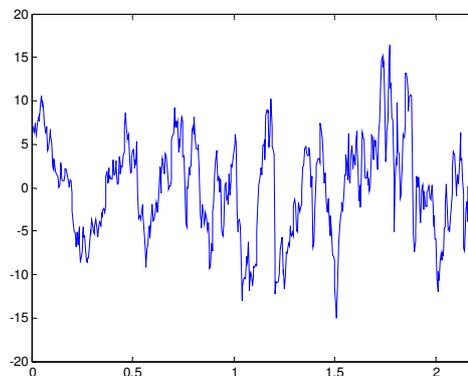


Figura 18 – Resíduo do S&P500 – 1987 do modelo log-periódico.

A fig.18 sugere a possibilidade de que ainda haja estrutura presente no resíduo que o modelo log-periódico não conseguiu capturar. É necessário então investigar as possíveis origens desta estrutura (não-estacionariedade, autocorrelação, heterocedasticidade, não-linearidades).

3.2.3. Estacionariedade do Resíduo

A estimação de modelos estatísticos em séries não estacionárias, como a série de preços dos índices financeiros, é potencialmente problemática, em consequência da possibilidade de ocorrência de regressões espúrias. (Hamilton, 1994, pg.557) [12]. Uma regressão é dito espúria quando seu resíduo for não estacionário para todos os valores possíveis dos parâmetros do modelo. Neste caso as estimativas obtidas pelos métodos de estimação não possuem sentido estatístico. Por exemplo, os estimadores não serão consistentes, ou seja, não convergem assintoticamente para os seus valores verdadeiros, fixos e desconhecidos.

Existem três soluções possíveis para se evitar a ocorrência de regressões espúrias:

- I. Incluir valor defasado da variável dependente e da variável independente na regressão;
- II. Trabalhar com a série na primeira diferença;
- III. Estimar o modelo em nível utilizando erro autoregressivo, tipicamente de ordem 1. Esta situação é assintoticamente equivalente à (II), para modelos lineares. (Hamilton, 1994, pg.562)

Para investigar a ocorrência de regressão espúria, avaliou-se a estacionariedade do resíduo do modelo log-periódico, através de estatísticas de teste de raiz unitária. A hipótese nula da estatística de teste é de raiz unitária (RU), ou seja, não-estacionariedade da série [11], [12].

Na tabela a seguir são apresentados os resultados das estatísticas de teste, utilizando os testes Phillips-Perron e ADF (Augmented Dickey-Fuller) disponíveis no programa Eviews V5.

Teste de Raiz Unitária		
	Sem intercepto	Com intercepto
Phillips-Perron	-4.919 (0.00)	-4.914 (0.00)
ADF	-5.208 (0.00)	-5.203 (0.00)

Tabela 3 - Teste de raiz unitária para o resíduo do S&P500 – 1987 do modelo log-periódico.

Da tabela 3, sendo os p-valores $\ll 0.05$, há evidência forte para rejeitar H_0 , ou seja, aceita-se a hipótese alternativa de que os resíduos do modelo log-periódico, aplicado ao nível da série, são estacionários.

3.2.4. Autocorrelação e Heterocedasticidade

Para investigar a presença de autocorrelação, analisa-se o correlograma do resíduo, onde na primeira coluna encontram-se os valores da função de autocorrelação (FAC), na segunda coluna os valores da função de autocorrelação parcial (PAC) e na terceira e quarta colunas os valores da estatística Q de Ljung-Box para o $lag(k)$ e o seu p-valor. A estatística Q de Ljung-Box, para o $lag(k)$, testa a hipótese nula de que não existe autocorrelação até a ordem k. O correlograma da série dos resíduos do modelo log-periódico para o S&P500 – 1987 é apresentado a seguir.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.922	0.922	464.74	0.000
		2	0.832	-0.119	843.80	0.000
		3	0.747	-0.008	1150.0	0.000
		4	0.663	-0.044	1392.2	0.000
		5	0.599	0.075	1589.6	0.000
		6	0.547	0.031	1754.9	0.000
		7	0.501	-0.002	1893.8	0.000
		8	0.459	-0.009	2010.4	0.000
		9	0.414	-0.039	2105.4	0.000
		10	0.367	-0.026	2180.3	0.000
		11	0.311	-0.083	2234.3	0.000
		12	0.262	0.016	2272.5	0.000
		13	0.215	-0.029	2298.3	0.000
		14	0.165	-0.064	2313.5	0.000
		15	0.128	0.045	2322.7	0.000
		16	0.105	0.041	2328.9	0.000
		17	0.086	-0.000	2333.0	0.000
		18	0.068	-0.021	2335.6	0.000
		19	0.052	0.001	2337.2	0.000
		20	0.037	0.008	2337.9	0.000

Verifica-se a presença de uma forte autocorrelação de 1ª ordem na FAC do resíduo do modelo log-periódico. Pelo correlograma observa-se que a estatística Q de Ljung-Box, para o $lag(20)$, possui p-valor $\ll 0.05$ ($Q(20) = 2337.9$ com p-valor = 0), sendo uma evidência forte de rejeição da hipótese nula de que não existe autocorrelação. Este resultado indica a necessidade de incorporar uma estrutura autoregressiva de ordem 1, AR(1), no modelo original [12].

Investigando também a presença de dependência a partir da FAC do quadrado do resíduo, u_t^2 , obtivemos os seguintes resultados:

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
			1	0.089	0.089	4.2980	0.038
			2	0.042	0.034	5.2562	0.072
			3	0.025	0.019	5.6032	0.133
			4	0.115	0.111	12.832	0.012
			5	0.119	0.100	20.585	0.001
			6	0.121	0.099	28.709	0.000
			7	0.060	0.035	30.667	0.000
			8	0.015	-0.011	30.800	0.000
			9	0.018	-0.010	30.985	0.000
			10	0.130	0.098	40.374	0.000
			11	0.112	0.070	47.417	0.000
			12	0.047	0.012	48.657	0.000
			13	-0.010	-0.030	48.709	0.000
			14	0.036	0.012	49.423	0.000
			15	0.083	0.046	53.327	0.000
			16	0.016	-0.037	53.478	0.000
			17	0.053	0.024	55.033	0.000
			18	0.001	-0.017	55.034	0.000
			19	0.029	0.017	55.496	0.000
			20	0.017	-0.008	55.656	0.000

Pelo correlograma observa-se que a estatística Q de Ljung-Box, para o $lag(20)$, possui p-valor $\ll 0.05$ ($Q(20) = 55.656$ com p-valor = 0), sendo uma evidência forte de rejeição da hipótese nula de que não existe autocorrelação na série do quadrado do resíduo.

Este resultado sugere assim a existência de dependência na variância condicional do resíduo do modelo log-periódico. A fim de capturar esta estrutura, deve ser incorporado ao modelo log-periódico um choque heterocedástico do tipo GARCH.

Assim, a análise destes resíduos, apesar de indicarem estacionariedade, indicam a presença de forte autocorrelação de 1ª ordem e de alguma heterocedasticidade condicional.

Somando as evidências apresentadas acima, foi, então, adotado um novo modelo, com a inclusão de um termo aleatório possuindo média condicional que evolui segundo um processo autoregressivo de ordem 1, AR(1), e variância condicional que evolui segundo um processo GARCH(1,1).¹⁰

¹⁰ O choque GARCH(1,1) foi escolhido por ser este um modelo bastante utilizado na prática dos mercados para descrever a heterocedasticidade condicional. Além disso, resultados empíricos de aplicações do modelo GARCH(1,1) indicam ser este suficiente para descrever a maioria das séries comumente analisadas no mercado.

3.3.

Estimação de Novo Modelo – Incorporando uma Estrutura AR e uma Estrutura GARCH

A segunda etapa do desenvolvimento do modelo proposto neste trabalho consiste em modificar a eq.(42), incorporando um processo AR(1) e um choque GARCH(1,1) ao resíduo u_t do modelo log-periódico original. Obtém-se assim, um novo modelo descrito por:

$$p_t = g(\Phi, t) + u_t \quad (44)$$

onde

$$g(\Phi, t) = A + B(t_c - t)^\beta + C(t_c - t)^\beta \cos[w \log(t_c - t) + \phi] \quad (45)$$

é a componente log-periódica e o resíduo u_t é dado por:

$$u_t = \rho u_{t-1} + \eta_t \quad (46)$$

$$\eta_t = \sigma_t \varepsilon_t \quad (47)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \eta_{t-1}^2 + \alpha_2 \sigma_{t-1}^2 \quad (48)$$

onde $\alpha_0 > 0$, $\alpha_i \geq 0$, $\alpha_1 + \alpha_2 < 1$ e $\varepsilon_t \sim N(0,1)$

Neste modelo, o resíduo u_t possui média condicional que evolui segundo um processo AR(1), dado por (46), e variância condicional σ_t^2 evoluindo segundo um processo GARCH(1,1) dada por (48). A média incondicional de u_t é zero e a variância incondicional de u_t é dada por:

$$\text{Var}[u_t] = \frac{\alpha_0}{(1 - \rho^2)(1 - \alpha_1 - \alpha_2)}.$$

Nesta etapa do procedimento, será estimado o vetor de parâmetros ótimos estendido $\hat{\Theta} = [A, B, C, t_c, \beta, w, \phi, \rho, \alpha_0, \alpha_1, \alpha_2]$.

A complexidade do modelo log-periódico original cresceu não linearmente com a incorporação das estruturas AR(1) e GARCH(1,1). Esta maior complexidade, acrescida da forte não-linearidade do modelo log-periódico levou a uma grande dificuldade na estimação da solução ótima. A estimação tornou-se mais onerosa, enfrentando problemas tais como não convergência ou convergência para mínimos locais. Para contornar estes problemas, a escolha de “boas” condições iniciais se tornou essencial.

O vetor das condições iniciais passou a ser construído em duas etapas, que estão ilustradas na figura 20 ao final desta seção:

1. As condições iniciais para $(A, B, C, t_c, \beta, w, \phi)$ são definidas como os parâmetros estimados para o modelo log-periódico original. Por construção do algoritmo do *Generalized Simulated Annealing*, garante-se que a solução obtida pelo GSA produz boas condições iniciais para $(A, B, C, t_c, \beta, w, \phi)$.
2. A partir da solução do GSA para o modelo log-periódico original, obtém-se o resíduo \hat{u}_t . Este resíduo, como já verificado na seção anterior, ainda possui estrutura de autocorrelação e heterocedasticidade. No programa Eviews, procede-se com a estimação simultânea de (46), (47) e (48), considerando \hat{u}_t um processo AR(1), com choque GARCH (1,1) na variância. Os parâmetros estimados pelo Eviews para o resíduo $\hat{u}_t : (\rho, \alpha_0, \alpha_1, \alpha_2)$, são utilizados como as condições iniciais para o modelo completo (44).

A função custo definida como a soma dos quadrados dos resíduos não atende ao modelo log-periódico - AR(1) - GARCH (1,1) devido à presença da heterocedasticidade. Um dos pressupostos para os estimadores de mínimos quadrados é de que $\text{Var}(u_t / p(t)) = E(u_t^2 / p(t)) = \sigma^2, \forall t$, ou seja, homocedasticidade.

Os estimadores dos parâmetros do modelo log-periódico - AR(1) - GARCH (1,1) são então obtidos pelo método da máxima verossimilhança condicional. Postulando $\varepsilon_t \sim N(0,1)$, a forma genérica do logaritmo da função de verossimilhança para o modelo GARCH (1,1) é dada por:

$$\ln L(\Theta) = -\frac{1}{2}(n-1)\ln(2\pi) - \frac{1}{2} \sum_{t=2}^n \ln \sigma_t^2 - \frac{1}{2} \sum_{t=2}^n \frac{\eta_t^2}{\sigma_t^2}$$

onde η_t^2 e σ_t^2 são dados respectivamente por (47) e (48).

Os resultados da estimação deste modelo aplicado à série do S&P500 estão na tabela a seguir. Para comparação, foram reproduzidos os resultados mostrados na tabela 1 para o modelo log-periódico original.

Crash SP500 – 1987		
	Log-per	Log-per-AR(1)-GARCH(1,1)
A	399.43	385.11
B	-153.06	-141.15
C	-12.09	-12.04
β	0.35	0.37
w	7.28	6.97
φ	1.19	1.41
ρ	-	0.935
α_0	-	0.023
α_1	-	0.036
α_2	-	0.962
t_c	2.239	2.210

Tabela 4 – Parâmetros estimados pelo modelo log-periódico original vs parâmetros estimados pelo modelo log-periódico - AR(1) - GARCH(1,1) do S&P500 – 1987.

Deve-se observar a pequena variação do valor dos parâmetros comuns aos dois modelos, mostrando a consistência da metodologia adotada.

O novo valor obtido para t_c também se mantém na faixa “economicamente” aceitável entre t_{max} e t_{min} , conforme mostrado a seguir.

Crash SP500 - 1987						
	t_c	Data t_c	t_{max}	Data t_{max}	t_{min}	Data t_{min}
Log-per	2.239	23-Sep-87	2.159	25-Aug-87	2.310	19-Oct-87
Log-per-AR(1) - GARCH(1,1)	2.210	11-Sep-87	2.159	25-Aug-87	2.310	19-Oct-87

Tabela 5 - Valor de t_c previsto pelos modelos e valores empíricos de $t_{máx}$ e $t_{mín}$.

Deve-se observar que no novo modelo t_c se aproxima de t_{max} , que caracteriza o primeiro dia do evento do *crash*, significando uma melhoria do poder preditivo da modelagem.

3.3.1. Investigação do Resíduo

A série temporal do resíduo padronizado $\varepsilon_t = \frac{\eta_t}{\sigma_t}$ deste modelo está mostrada no gráfico abaixo.

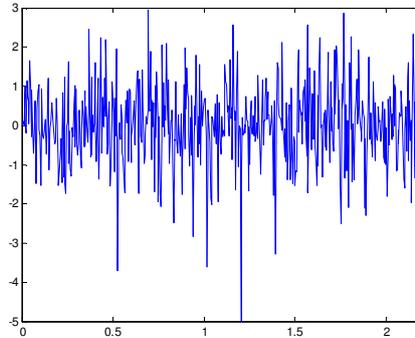


Figura 19 - Resíduo padronizado do S&P500 – 1987 do modelo log-periódico - AR(1) - GARCH(1,1).

Analisando a presença de autocorrelação na FAC do resíduo $\hat{\varepsilon}_t$ padronizado do modelo log-periódico - AR(1) - GARCH(1,1), obtivemos os seguintes resultados:

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
			1	0.089	0.089	4.3591	0.037
			2	-0.047	-0.056	5.5753	0.062
			3	0.023	0.033	5.8605	0.119
			4	-0.025	-0.034	6.2147	0.184
			5	-0.040	-0.032	7.1042	0.213
			6	-0.042	-0.039	8.0666	0.233
			7	-0.004	0.002	8.0748	0.326
			8	0.027	0.025	8.4918	0.387
			9	0.026	0.022	8.8761	0.449
			10	0.059	0.055	10.814	0.372
			11	-0.029	-0.042	11.278	0.420
			12	-0.005	0.007	11.290	0.504
			13	0.028	0.024	11.729	0.550
			14	-0.076	-0.074	14.959	0.381
			15	-0.065	-0.046	17.362	0.298
			16	0.008	0.011	17.400	0.360
			17	-0.030	-0.036	17.891	0.396
			18	-0.015	-0.011	18.022	0.454
			19	-0.001	-0.008	18.023	0.521
			20	-0.001	-0.011	18.023	0.586

A estatística Q de Ljung-Box, para o lag(20), possui p-valor $\gg 0.05$ (Q(20) = 18.023 com p-valor = 0.586), sendo uma evidência forte para aceitar a

hipótese nula de que não existe autocorrelação na série do resíduo do modelo log-periódico - AR(1) - GARCH(1,1).

Finalmente, analisando a presença de dependência na variância condicional a partir da FAC do quadrado do resíduo padronizado $\hat{\varepsilon}_t^2$ do modelo log-periódico - AR(1) - GARCH(1,1), obtém-se:

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.009	0.009	0.0469	0.829
		2	-0.041	-0.042	0.9888	0.610
		3	-0.024	-0.023	1.3010	0.729
		4	0.008	0.006	1.3326	0.856
		5	0.014	0.012	1.4433	0.920
		6	0.011	0.011	1.5081	0.959
		7	-0.014	-0.013	1.6220	0.978
		8	-0.027	-0.025	2.0192	0.980
		9	0.019	0.019	2.2176	0.988
		10	0.047	0.044	3.4463	0.969
		11	0.087	0.087	7.6308	0.746
		12	-0.029	-0.025	8.1020	0.777
		13	-0.041	-0.032	9.0534	0.769
		14	-0.010	-0.009	9.1110	0.824
		15	0.058	0.052	10.993	0.753
		16	0.041	0.037	11.948	0.748
		17	-0.023	-0.019	12.255	0.784
		18	-0.045	-0.036	13.422	0.766
		19	-0.002	0.000	13.425	0.816
		20	-0.028	-0.041	13.884	0.836

A estatística Q de Ljung-Box, para o $lag(20)$, possui p-valor $\gg 0.05$ ($Q(20) = 13.884$ com p-valor = 0.836), sendo uma evidência forte para aceitar a hipótese nula de que não existe autocorrelação na série do quadrado do resíduo do modelo log-periódico - AR(1) - GARCH(1,1).

Assim, nos correlogramas acima da FAC do resíduo padronizado e da FAC do quadrado do resíduo padronizado do modelo log-periódico - AR(1) - GARCH(1,1), a estatística Q de Ljung-Box, para o $lag(20)$, aceita a hipótese nula de que não existe autocorrelação até a ordem 20. Este resultado indica que este modelo foi capaz de capturar a autocorrelação e a heterocedasticidade que estavam presentes no resíduo do modelo log-periódico original.

A figura 20 adiante apresenta o esquema utilizado na estimação dos parâmetros do modelo log-periódico - AR(1) - GARCH (1,1).

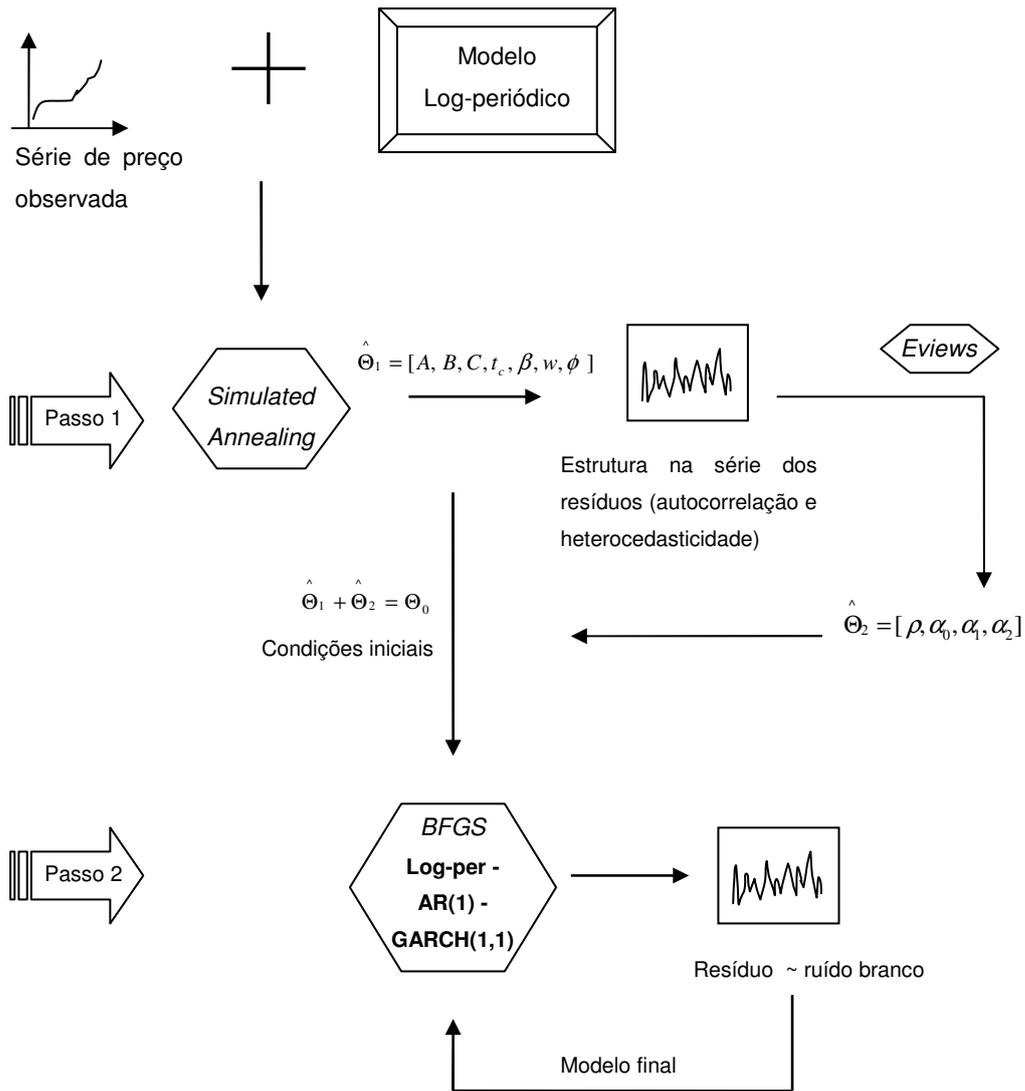


Figura 20 – Esquema do procedimento de estimação do modelo log-periódico – AR(1) – GARCH(1,1).

3.3.2. Estatísticas Descritivas e Teste BDS

Prosseguindo com a investigação da série do resíduo do modelo log-periódico - AR(1) - GARCH(1,1), foram avaliadas as estatísticas descritivas do resíduo, incluindo teste de normalidade (Jarque-Bera (JB) e QQ-plot), o histograma e, finalmente, procedendo com o teste BDS [13].

O QQ-plot é um procedimento que representa de forma “linearizada” a informação do gráfico da distribuição empírica acumulada das observações. Sob a hipótese de normalidade, se a distribuição teórica (Normal padrão) for uma boa aproximação para a distribuição empírica, então os pontos do gráfico do QQ-plot

descreverão uma reta que passa pela origem formando um ângulo de 45° com a horizontal.

A estatística BDS testa a hipótese nula de uma série temporal ser independente e identicamente distribuída (iid) contra hipóteses alternativas de dependência linear, ou dependência não linear, ou ainda, caos determinístico.

Os resultados das estatísticas estão mostrados a seguir.

Crash SP500 - 1987	
Média	0.017
Mediana	0.034
Máximo	2.953
Mínimo	-4.984
Desvio padrão	0.988
Assimetria	-0.384
Curtose	4.774
Jarque-Bera	84.673 (0.00)

Tabela 6 – Estatísticas descritivas e teste JB do resíduo do modelo log-periódico - AR(1) - GARCH(1,1) para o S&P500 – 1987.

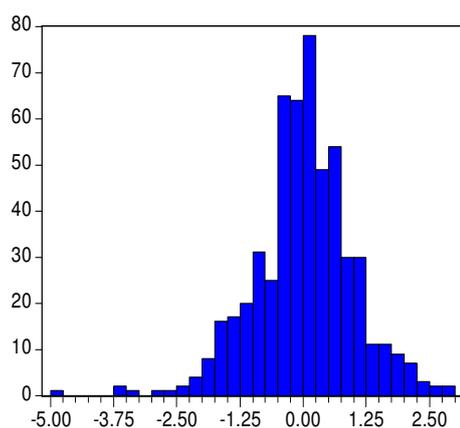


Figura 21 - Histograma do resíduo do S&P500 – 1987.

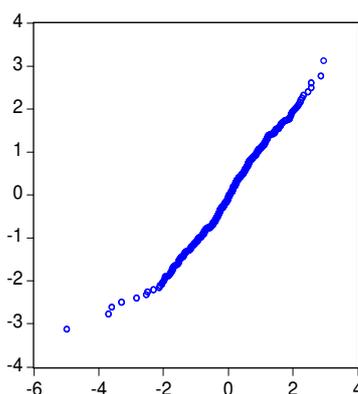


Figura 22 - QQ plot do resíduo do S&P500 – 1987.

m \ ϵ	0.5σ	1.0σ	1.5σ	2.0σ
2	0.92120	0.78320	0.75120	0.85120
3	0.70200	0.94880	0.97040	0.86440
4	0.79320	0.96400	0.99160	0.76360
5	0.57560	0.92400	0.87200	0.47920
6	0.57240	0.90920	0.98160	0.66520

Tabela 7 – p-valores *bootstrapeados* da estatística BDS, com 5000 repetições.

A partir da tabela 6, verifica-se que a distribuição da série do resíduo padronizado do modelo log-periódico - AR(1) - GARCH(1,1) apresenta adequadamente média aproximadamente 0 e variância aproximadamente 1. No

entanto, a hipótese de normalidade é fortemente rejeitada pelo teste Jarque-Bera (p-valor $\ll 0.05$), o que é de esperar já que a distribuição é leptocúrtica, ou seja, possui curtose maior do que 3 (sendo 3 o valor de curtose da Normal padrão). Da mesma forma, o histograma e o QQ-plot, mostrados respectivamente nas figuras 21 e 22, confirmam a não-normalidade da série do resíduo padronizado.

O histograma indica visualmente a presença de curtose, que é quantificada na tabela 6. A presença de curtose é considerada natural para o período analisado predecessor ao *crash*, onde é esperada maior frequência de ocorrência de grandes flutuações nos preços dos índices financeiros quando comparada com o regime normal do mercado.

A distribuição da estatística BDS sob a hipótese nula de iid é assintoticamente Normal, de média 0 e variância 1. No entanto, para pequenas amostras ou para séries com distribuição não usual, como é o caso da série de resíduo padronizado do S&P500 do modelo log-periódico - AR(1) - GARCH(1,1), não é garantido que a distribuição da estatística BDS irá tender assintoticamente para uma Normal padrão. Desta forma, a estatística de teste BDS deve ser levantada empiricamente através de *bootstrap*.

Assim, na tabela 7, são apresentados os p-valores do teste BDS, obtidos a partir de *bootstrap* com 5000 repetições. A hipótese nula é fortemente aceita¹¹, ou seja, a série do resíduo padronizado do modelo log-periódico - AR(1) - GARCH(1,1) é independente e identicamente distribuída (iid).

3.4. Inferência em Modelo com Tendência Não-linear

Aceitando que os resíduos sejam estacionários, como foi visto na seção 3.2.3, a questão agora é sobre a inferência possível em um modelo com tendência determinística não linear, como é o caso do modelo log-periódico, descrito pelo processo gerador p_t , eq.(42).

Sabe-se que para a classe de processos com tendência determinística os estimadores de MQO não possuem as propriedades padrões observadas nos modelos de regressão envolvendo variáveis estacionárias, apresentando em geral diferentes taxas de convergência e expressões para a variância assintótica.

¹¹ No teste BDS se aceita a hipótese nula apenas quando os p-valores forem maiores ou iguais a 0.05 para **todos** os elementos da matriz $m \setminus \varepsilon$ (tabela 7).

No Hamilton, 1994, pg. 454 [12], são apresentados os resultados teóricos para estimação de MQO para um processo com tendência linear determinística e choque ruído branco. O resultado principal é que as estatísticas de teste usuais t e F neste caso possuem as mesmas distribuições assintóticas encontradas para processos estacionários, embora a taxa de convergência para estas distribuições seja diferente da usual: ao invés de convergência em \sqrt{T} converge em $T^{3/2}$, sendo T o tamanho da série das observações. Na literatura econométrica no entanto não existem resultados semelhantes quando se trata de funções não lineares e determinísticas do tempo.

Como a obtenção de estatísticas adequadas para uma correta inferência destes modelos está fora do escopo desta dissertação, neste trabalho será conjecturado que os resultados encontrados no Hamilton, 1994, pg. 454, podem ser estendidos para as inferências no modelo log-periódico. Assim sendo, será suposto que as inferências para o modelo log-periódico serão confiáveis dentro dos mesmos critérios.