Inovações Estáveis em Processos ARFIMA(p,d,q)

Gennaro Anesi e Sílvia R.C. Lopes
Instituto de Matemática - Universidade Federal do Rio Grande do Sul
gennaroanesi@gmail.com e silvia.lopes@ufrgs.br

28 de Abril de 2010

Resumo

Neste trabalho é apresentado um estudo sobre os processos ARFIMA(p,d,q) com decaimento hiperbólico da função de autocorrelação. Foram analisadas séries temporais com inovações estáveis, isto é, distribuições caracterizadas pela presença de caudas pesadas. O objetivo principal é verificar a consistência de diversos estimadores utilizados para o parâmetro de diferenciação d do processo e para o parâmetro de estabilidade α das distribuições estáveis.

Palavras-chave: ARFIMA; Distribuições Estáveis; Séries Temporais; Processos com Longa Dependência.

1 Introdução

No estudo de séries temporais, há a necessidade de modelos que possam apresentar a característica de longa dependência, isto é, mesmo para observações separadas por um longo período de tempo, a correlação verificada não é desprezível. Esta particularidade tem sido observada em séries temporais de diferentes áreas, tais como meteorologia, hidrologia e economia. Estudos de séries que apresentam esta característica foram feitos inicialmente por Hurst (1951). De fato, um destes modelos é o auto-regressivo média móvel fracionariamente integrável, denotado por ARFIMA(p,d,q). Em geral, ele possui como inovações um processo ruído branco. Contudo, neste trabalho, foram analisados processos cujas inovações seguem distribuições estáveis. É intrínseca a estas distribuições a presença de caudas pesadas, o que ocasiona o aparecimento de valores extremos nas séries temporais. Esta característica não ocorre quando as inovações são Gaussianas.

Neste contexto, objetiva-se estimar os parâmetros de diferenciação d da série temporal (que determina se a série possui a característica de longa dependência) e α das distribuições estáveis. Uma vez estimado o parâmetro d, diferencia-se a série temporal de $\hat{d}$ e então estima-se o parâmetro α das inovações do processo. Para tal objetivo, foram utilizados vários estimadores, tanto da classe paramétrica quanto da classe semi-paramétrica.

2 Distribuições α-estáveis

Descobertas por Lévy (1924), as distribuições estáveis possuem uma vasta área de aplicações como, por exemplo, teoria das probabilidades, teoria das comunicações, física, astronomia, economia, sociologia, etc. A seguir, algumas definições fundamentais utilizadas neste trabalho.

Definição 2.1 (Distribuições Estáveis)**.** Uma variável aleatória (v.a.) X ou sua função de distribuição (f.d.) F é estável se e somente se para quaisquer números positivos A e B existem números positivos C e D tais que

$$AX_1 + BX_2 \stackrel{d}{=} CX + D, \tag{1}$$

onde X_1 e X_2 são cópias independentes de X.

Proposição 2.1 (Teorema do Limite Central Generalizado)**.** *Sejam $X_1, X_2, ...$ v.a.'s independentes e identicamente distribuídas (i.i.d.). Uma v.a. X é estável se existem $a_n > 0$ e b_n tais que*

$$\frac{S_n - b_n}{a_n} \xrightarrow{d} X, \text{ quando } n \longrightarrow \infty,$$

onde $S_n = X_1 + X_2 + \ldots + X_n$ e $\xrightarrow{d}$ denota convergência em distribuição.

Um novo método de integração numérica de duas quadraturas foi utilizado para a obtenção da densidade de probabilidade, visando resolver os problemas de oscilação nas caudas observados em métodos tradicionais, conforme se verifica em Belov (2005).

3 Modelo ARFIMA(p,d,q)

O modelo auto-regressivo média móvel fracionariamente integrável, denotado por ARFIMA(p, d, q), é capaz de descrever a característica de *longa dependência* se $d \in$(0,0;0,5). Esta característica tem sido observada em séries temporais de diferentes áreas tais como meteorologia, hidrologia e economia. Veja Beran (1994) e Doukhan et al. (2003) para um estudo completo destes modelos.

Definição 3.1 (Processo ARFIMA)**.** Seja $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ um processo estocástico cujas variáveis aleatórias são independentes e identicamente distribuídas. Seja $\mathcal{B}$ o operador de defasagem, i.e.,

$$\mathcal{B}^j(X_t) = \mathcal{B}^{j-1}(\mathcal{B}(X_t)) = \mathcal{B}^{j-1}(X_{t-1}) = X_{t-j}.$$

Se $\{X_t\}_{t\in\mathbb{Z}}$ é um processo linear satisfazendo a equação

$$\Phi(\mathcal{B})(1-\mathcal{B})^d(X_t - \mu) = \Theta(\mathcal{B})\varepsilon_t, \; t \in \mathbb{Z}, \tag{2}$$

onde $d \in (-0,5; 0,5)$, $\Phi(\cdot)\, e\, \Theta(\cdot)$ são os polinômios de ordem p e q, respectivamente, dados por

$$\Phi(\mathcal{B}) = 1 - \phi_1\mathcal{B} - \cdots - \phi_{\mathrm{p}}\mathcal{B}^{\mathrm{p}}$$

$$\Theta(\mathcal{B}) = 1 - \theta_1\mathcal{B} - \cdots - \theta_{\mathrm{q}}\mathcal{B}^{\mathrm{q}},$$

onde ϕ_i, $1 \leqslant i \leqslant p$, e θ_j, $1 \leqslant j \leqslant q$, são constantes reais com $\phi_0 = 1 = \theta_0$, e μ é a sua média, então é chamado de *processo geral com diferenciação fracionária* ARFIMA(p,d,q), onde d é o *grau* ou *parâmetro de diferenciação.*

4 Estimação do Parâmetro de Longa Dependência

Para estimar o parâmetro de diferenciação fracionário d dos processos estocásticos ARFIMA (p,d,q) foram utilizados estimadores paramétricos e semi-paramétricos, clássicos e robustos. São eles: MLE ("Maximum Likelihood Estimator"), proposto por Sowell (1992); LW ("Local Whittle"), proposto por Robinson (1995); BALM, BAMM e BALTS, proposto por Lopes e Mendes (2006); FELW ("Fully Extended Local Whittle"), proposto por Abadir et al. (2007); DFA ("Detrended Fluctuation Analysis"), proposto por Peng et al. (1994); W ("Whittle"), proposto por Fox e Taqqu (1986); GPHLM proposto por Geweke e Porter-Hudak (1983) e suas versões robustas, GPHMM e GPHLTS; e GSE ("Gaussian Semiparametric Estimator"), proposto por Robinson (1995) e aperfeiçoado por Lobato (1999) e Shimotsu (2007).

5 Estimação do Parâmetro α das Inovações Estáveis

Após estimar o parâmetro de diferenciação fracionário d das séries temporais, deve-se diferenciar cada uma delas para então obter seu ruído (ou o que estimamos que seja seu ruído). Em seguida, estima-se o parâmetro α deste ruído, que deve seguir uma distribuição α-estável. Para tanto, utiliza-se o estimador de máxima verossimilhança de Hill-Hall, denotado neste por HH (ver Crato, 2000). Nas tabelas da seção 6.2, é destacado o ponto de truncamento r utilizado na estimação deste parâmetro.

6 Simulações

Esta seção consiste na apresentação de algumas simulações realizadas. Foram simuladas séries temporais seguindo o modelo ARFIMA(0,d,0) com $d \in$(0;0,5); as inovações seguiram distribuições estáveis com $\alpha \in$(1;2). A seguir, estimou-se o valor de d em cada uma delas a partir dos estimadores supracitados e, diferenciando as séries a partir destes valores estimados de d, pôde-se estimar então os valores de α em cada uma das séries. Todas as simulações consistiram em 1000 replicações de cada tipo de série, com séries de 500 observações. Os pontos de truncamento usados foram: $g(n) = n^{0,85}$ para os estimadores FELW, GPHLM, GPHMM, GPHLTS, BALM, BAMM e BALTS, e $g(n) = n^{0,5}$ para o DFA (ver Tabela 1). Para o estimador de α, utilizamos diversos pontos de truncamento r, como mostram as Tabelas 2 e 3.

6.1 Estimação do parâmetro d

Tabela 1: Estimação de d

Estimador	MLE	LW	FELW	DFA	BALM	BAMM	BALTS	W	GPHLM	GPHMM	GPHLTS	GSE
						$d = 0,2$ e $\alpha = 1,25$						
Média	0.1913	0.2027	0.1960	0.1967	0.1989	0.2067	0.2328	0.1912	0.1913	0.2027	0.1960	0.2034
Vício	-0.0087	0.0027	-0.0040	-0.0033	-0.0011	0.0067	0.0328	-0.0088	-0.0087	0.0027	-0.0040	0.0034
eqm	0.0012	0.0041	0.0014	0.0050	0.0019	0.0048	0.0183	0.0013	0.0012	0.0041	0.0014	0.0459
Variância	0.0011	0.0041	0.0014	0.0050	0.0019	0.0048	0.0173	0.0012	0.0011	0.0041	0.0014	0.0045
						$d = 0,2$ e $\alpha = 1,5$						
Média	0.1910	0.1954	0.1940	0.1965	0.1964	0.2118	0.2255	0.1901	0.1910	0.1954	0.1940	0.1963
Vício	-0.0090	-0.0046	-0.0060	-0.0035	-0.0036	0.0118	0.0255	-0.0099	-0.0090	-0.0046	-0.0060	-0.0037
eqm	0.0017	0.0046	0.0016	0.0022	0.0020	0.0035	0.0151	0.0014	0.0017	0.0046	0.0016	0.0434
Variância	0.0016	0.0046	0.0016	0.0022	0.0020	0.0034	0.0144	0.0013	0.0016	0.0046	0.0016	0.0049
						$d = 0,2$ e $\alpha = 1,75$						
Média	0.1909	0.1994	0.1952	0.1971	0.1991	0.2253	0.2232	0.1903	0.1909	0.1994	0.1952	0.1976
Vício	-0.0091	-0.0006	-0.0048	-0.0029	-0.0009	0.0253	0.0232	-0.0097	-0.0091	-0.0006	-0.0048	-0.0024
eqm	0.0013	0.0052	0.0017	0.0032	0.0022	0.0044	0.0070	0.0014	0.0013	0.0052	0.0017	0.0447
Variância	0.0013	0.0052	0.0017	0.0032	0.0022	0.0038	0.0065	0.0013	0.0013	0.0052	0.0017	0.0057
						$d = 0,3$ e $\alpha = 1,5$						
Média	0.2911	0.3021	0.2980	0.2995	0.3002	0.2988	0.3080	0.2938	0.2911	0.3021	0.2980	0.3020
Vício	-0.0089	0.0021	-0.0020	-0.0005	0.0002	-0.0012	0.0080	-0.0062	-0.0089	0.0021	-0.0020	0.0020
eqm	0.0014	0.0050	0.0017	0.0024	0.0022	0.0031	0.0119	0.0014	0.0014	0.0050	0.0017	0.0965
Variância	0.0013	0.0050	0.0017	0.0024	0.0022	0.0031	0.0119	0.0014	0.0013	0.0050	0.0017	0.0053
						$d = 0,3$ e $\alpha = 1,75$						
Média	0.2886	0.2971	0.2958	0.2983	0.2993	0.3033	0.3039	0.2909	0.2886	0.2971	0.2958	0.2958
Vício	-0.0114	-0.0029	-0.0042	-0.0017	-0.0007	0.0033	0.0039	-0.0091	-0.0114	-0.0029	-0.0042	-0.0042
eqm	0.0013	0.0048	0.0017	0.0026	0.0022	0.0028	0.0058	0.0014	0.0013	0.0048	0.0017	0.0927
Variância	0.0012	0.0048	0.0017	0.0026	0.0022	0.0027	0.0058	0.0013	0.0012	0.0048	0.0017	0.0052
						$d = 0,4$ e $\alpha = 1,75$						
Média	0.3869	0.4012	0.3988	0.4008	0.4020	0.3881	0.3896	0.3928	0.3869	0.4012	0.3988	0.4027
Vício	-0.0131	0.0012	-0.0012	0.0008	0.0020	-0.0119	-0.0104	-0.0072	-0.0131	0.0012	-0.0012	0.0027
eqm	0.0014	0.0051	0.0018	0.0026	0.0023	0.0027	0.0056	0.0014	0.0014	0.0051	0.0018	0.1678
Variância	0.0013	0.0051	0.0018	0.0026	0.0023	0.0026	0.0055	0.0013	0.0013	0.0051	0.0018	0.0056

6.2 Estimação de α Através do Estimador de Máxima Verossimilhança de Hill-Hall

Tabela 2: Estimação de α quando $d = 0,2$ e $\alpha = 1,25$

Estimador	MLE	LW	BALM	BAMM	BALTS	FELW	DFA	W	GPHLM	GPHMM	GPHLTS	GSE
						$r = 0.65$						
Média	1.3412	1.3485	1.3442	1.3494	1.3412	1.3478	1.3204	1.3409	1.5027	1.5027	1.4412	1.3524
Vício	0.0912	0.0985	0.0942	0.0994	0.0912	0.0978	0.0704	0.0909	0.2527	0.2527	0.1912	0.1024
eqm	0.0872	0.0904	0.0851	0.1302	0.0852	0.0951	0.0847	0.0868	0.3745	0.3745	0.2618	0.0916
Variância	0.0789	0.0808	0.0763	0.1205	0.0770	0.0856	0.0799	0.0786	0.3110	0.3110	0.2254	0.0812
						$r = 0.7$						
Média	1.3093	1.3165	1.3150	1.3093	1.3102	1.3157	1.2982	1.3098	1.4722	1.4722	1.4308	1.3188
Vício	0.0593	0.0665	0.0650	0.0593	0.0602	0.0657	0.0482	0.0598	0.2222	0.2222	0.1808	0.0688
eqm	0.0837	0.0844	0.0784	0.1320	0.0848	0.0922	0.0904	0.0829	0.4185	0.4185	0.3217	0.0883
Variância	0.0803	0.0800	0.0743	0.1286	0.0813	0.0880	0.0881	0.0794	0.3696	0.3696	0.2893	0.0837
						$r = 0.75$						
Média	1.2330	1.2343	1.2334	1.2274	1.2347	1.2391	1.2343	1.2330	1.4055	1.4055	1.3790	1.2389
Vício	-0.0170	-0.0157	-0.0166	-0.0226	-0.0153	-0.0109	-0.0157	-0.0170	0.1555	0.1555	0.1290	-0.0111
eqm	0.0837	0.0893	0.0791	0.1319	0.0841	0.0947	0.0966	0.0836	0.4315	0.4315	0.3598	0.0968
Variância	0.0835	0.0892	0.0789	0.1315	0.0839	0.0947	0.0964	0.0834	0.4078	0.4078	0.3435	0.0968
						$r = 0.8$						
Média	1.0766	1.0800	1.0811	1.0657	1.0775	1.0802	1.0842	1.0781	1.2899	1.2899	1.2739	1.0803
Vício	-0.1734	-0.1700	-0.1689	-0.1843	-0.1725	-0.1698	-0.1658	-0.1719	0.0399	0.0399	0.0239	-0.1697
eqm	0.1263	0.1235	0.1166	0.1524	0.1239	0.1312	0.1370	0.1246	0.4630	0.4630	0.4109	0.1321
Variância	0.0963	0.0947	0.0882	0.1186	0.0942	0.1025	0.1096	0.0951	0.4619	0.4619	0.4108	0.1034

Tabela 3: Estimação de α quando $d = 0,3$ e $\alpha = 1,5$

Estimador	MLE	LW	BALM	BAMM	BALTS	FELW	DFA	W	GPHLM	GPHMM	GPHLTS	GSE
						$r = 0.65$						
Média	1.8229	1.8253	1.8244	1.8268	1.8255	1.8294	1.7860	1.8219	2.0226	2.0226	1.9122	1.8247
Vício	0.3229	0.3253	0.3244	0.3268	0.3255	0.3294	0.2860	0.3219	0.5226	0.5226	0.4122	0.3247
eqm	0.2048	0.2017	0.2010	0.2879	0.2019	0.2041	0.1690	0.2014	1.1382	1.1382	0.6742	0.2070
Variância	0.1006	0.0960	0.0959	0.1813	0.0961	0.0957	0.0873	0.0979	0.8660	0.8660	0.5048	0.1017
						$r = 0.7$						
Média	1.7895	1.7826	1.7871	1.7819	1.7869	1.7914	1.7680	1.7877	1.9165	1.9165	1.8463	1.7815
Vício	0.2895	0.2826	0.2871	0.2819	0.2869	0.2914	0.2680	0.2877	0.4165	0.4165	0.3463	0.2815
eqm	0.1656	0.1662	0.1617	0.2453	0.1613	0.1660	0.1483	0.1633	1.0031	1.0031	0.6720	0.1663
Variância	0.0819	0.0865	0.0793	0.1660	0.0791	0.0812	0.0766	0.0806	0.8305	0.8305	0.5527	0.0871
						$r = 0.75$						
Média	1.6477	1.6405	1.6453	1.6432	1.6512	1.6556	1.6392	1.6488	1.7585	1.7585	1.7158	1.6380
Vício	0.1477	0.1405	0.1453	0.1432	0.1512	0.1556	0.1392	0.1488	0.2585	0.2585	0.2158	0.1380
eqm	0.0870	0.0908	0.0835	0.1502	0.0841	0.0878	0.0846	0.0855	0.8650	0.8650	0.6387	0.0960
Variância	0.0652	0.0711	0.0625	0.1298	0.0613	0.0637	0.0653	0.0634	0.7990	0.7990	0.5928	0.0770
						$r = 0.8$						
Média	1.3526	1.3422	1.3492	1.3473	1.3505	1.3531	1.3524	1.3522	1.5518	1.5518	1.5273	1.3456
Vício	-0.1474	-0.1578	-0.1508	-0.1527	-0.1495	-0.1469	-0.1476	-0.1478	0.0518	0.0518	0.0273	-0.1544
eqm	0.0831	0.0886	0.0834	0.1308	0.0823	0.0835	0.0851	0.0824	0.7675	0.7675	0.6214	0.0906
Variância	0.0614	0.0638	0.0607	0.1076	0.0600	0.0620	0.0634	0.0606	0.7658	0.7658	0.6214	0.0668

7 Conclusão

Neste estudo, verificou-se boa consistência na estimação de ambos os parâmetros. Na estimação do parêmetro de longa dependência d do processo ARFIMA(p, d, q), observou-se que todos os estimadores apresentaram pouco vício, ou seja, o valor estimado mostrou-se próximo do valor real. Notou-se ainda que a variância dos estimadores foi pequena, o que verifica sua consistência. Já para a estimação do parâmetro α das distribuições estáveis, evidencia-se a necessidade da escolha de um bom ponto de truncamento para o estimador de Hill-Hall. De fato, neste trabalho, notou-se que os melhores resultados foram obtidos quando $r \in$(1;2) e, ainda, que quanto maior o valor de d, maior deve ser o ponto de truncamento utilizado para que a estimativa de α seja melhor.

Agradecimentos

Gennaro Anesi agradece o suporte do CNPq e da UFRGS; Sílvia R.C. Lopes agradece o suporte parcial do CNPq, da CAPES, do INCT em *Matemática* e também do Pronex *Probabilidade e Processos Estocásticos* -E-26/170.008/2008 -APQ1.

8 Referências

Abadir, K.M., W. Distaso e L. Giraitis (2007). "Nonstationarity- extended local Whittle estimation". *Journal of Econometrics*, Vol. **141**, 1353-1384.

Belov, I.A. (2005). "On the computation of the probability density function of α-stable distributions". *Mathematical Modelling and Analysis 2005*, 333-341.

Beran, J. (1994). "Statistics for Long-Memory Processes". New York: Chapman & Hall.

Crato, N. (2000). "Estimation of the maximal moment exponent with censored data". *Communications in Statistics. Simulation and Computation*, Vol. **29**(4), 239-1253.

Fox, R. e Taqqu, M.S. (1986). "Large sample properties of parameter estimates for strongly dependent stationary time series". *Annals of Statistics*, Vol. **14**, 517-532.

Geweke, J. e Porter-Hudak, S. (1983). "The estimation and application of long memory time series models". *Journal of Times Series Analysis*, Vol. **4**, 221-238.

Hurst, H.E. (1951). "Long-term storage capacity of reservoirs". *Transactions of the American Society of Civil Engineers*. Vol. **116**, 770–799.

Lévy, P. (1924). "Théorie des erreurs la loi de Gauss et les lois exceptionelles" ["Theory of errors: The Law of Gauss and the exceptional laws"] . *Bulletin de la Société Mathématique de France*, Vol. **52**, 49-85.

Ling, S. e Peng, L. (2004). "Hill's estimator for the tail index of an ARMA model". *Journal of Statistical Planning and Inference*, Vol. **123**, 279-293.

Lobato, I.N. (1999). "A semiparametric two-step estimator in a multivariate long memory model". *Journal of Econometrics*, Vol **90**, 129-153.

Lopes, S.R.C. e B.V.M. Mendes (2006). "Bandwidth Selection in Classical and Robust Estimation of Long Memory". *International Journal of Statistics and Systems*, Vol. **1**(1), 167-190.

Robinson, P.M. (1995a). "Log-Periodogram Regression of Time Series with Long Range Dependence". *Annals of Statistics*, Vol. **23**(3), 1048-1072.

Robinson, P.M. (1995b). "Robinson Gaussian Semiparametric Estimation of Long Range Dependence". *The Annals of Statistics*, Vol. **23**(5), 1630-1661.

Samorodnitsky, G. e Taqqu, M.S. (1994). "Stable Non-Gaussian Random Processes". New York: Chapman & Hall.

Shimotsu, K. (2007). "Gaussian semiparametric estimation of multivariate fractionally integrated processes". *Journal of Econometrics*, Vol. **137**(2), 277-310.

Sowell, F. (1992). "Maximum Likelihood Estimation of Stationary Univariate Fractionally Integrated Time Series Models". *Journal of Econometrics*, Vol. **53**, 165-188.

Whittle, P. (1953). "Estimation and information in stationary time series". *Arkiv for Matematik*, Vol. **2**, 423-434.