

ESTIMAÇÃO ROBUSTA DE PROCESSOS ARFIMA

FABIO ALEXANDER FAJARDO MOLINARES

Orientador: Prof. Valdério A. Reisen, Ph.D.

Area de Concentração: Séries Temporais

Tese submetida como requerimento parcial para obtenção do grau de Doutor em Estatística pela Universidade Federal de Minas Gerais

Belo Horizonte, Agosto de 2011

*A mi esposa Nátaly y mi hija Sara, a mis padres Carmen y Jorge, a mis hermanos
Olga, Edgar, Marlon y Jorge, a mis sobrinos Danna, Edgar, Karla, Mariangel y
Katherine, a mi amigototota de toda la vida Keyla.*

AGRADECIMENTOS

À professora Ela Mercedes Medrano de Toscano, pela amizade, ajuda invaluable, constante apoio e pelas intermináveis discussões que se tornaram um grande incentivo para o desenvolvimento deste trabalho, contribuindo para o meu amadurecimento acadêmico e profissional. A ela dedico todo o esforço que investi no desenvolvimento deste trabalho, assim como as futuras publicações que resultem como consequência do mesmo.

A Nátaly e Sara, meus grandes amores, por todos os momentos de alegria, pelo constante apoio e pelo incentivo para que todas as metas traçadas fossem alcançadas.

Ao professor Valdério A. Reisen pela amizade e apoio incondicional durante o mestrado e doutorado.

A César Fajardo, Zuleima, César Felipe, John Sebastian, Clementina e Nieves Camacho, pelo carinho e pelo constante apoio.

A meus compadres Tei e Wil, a meus sogros Teresa e Salvador, a minha cunhada Gigi. A todos eles pelo infinito apoio, ótimos conselhos, amizade e incentivo.

A Rogeria Ferreira, pela presteza, constante apoio e amabilidade com que sempre me ajudou.

Aos meus amigos na Colômbia, em especial a Camilo e Pacho.

Aos professores e amigos Rafael E. Ahumada Barrios e Francisco J. Cepeda Coronado, pelos

ótimos conselhos e pela orientação segura na minha vida acadêmica e profissional.

Ao meu padrinho Luis A. López, pela amizade e pela contribuição à formação profissional.

Aos meus amigos no Brasil, em especial a Lúcia Golçalves, Alyne Neves, Bartolomeu Zamprogno, Jacqueline Alves, Alessandro Sarnaglia, Giovanni Comarela, Marcelo Bourguignon e Lia Reisen em Vitória, Mario Piscoya e Rivana Alves em Belo Horizonte, Márcia H. Barbian em Porto Alegre, pela amizade, apoio emocional e momentos de diversão.

Aos meus colegas do Mestrado e Doutorado em Estatística da Universidade Federal de Minas Gerais, pelas horas compartilhadas.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, pelo apoio financeiro.

RESUMO

Este documento concentra-se no estudo de metodologias usadas para estimação do parâmetro de memória em processos ARFIMA contaminados com dados atípicos. As metodologias sugeridas são baseadas em estimadores robustos do espectro do processo; e apresentam-se como alternativas razoáveis para estimação do parâmetro de diferenciação fracionária em processos contaminados. Os estimadores robustos do espectro são classificados em duas classes: estimadores tipo M e estimadores tipo L . Na primeira classe o estimador robusto do espectro é função da soma de quadrados dos coeficientes de Fourier do processo; enquanto que na segunda o estimador requer o cálculo de um estimador robusto da função de autocovariância, baseada em L -estimadores (ver, Ma & Genton (2000)). O cálculo de estimadores robustos para o espectro do processo, permite o desenvolvimento de propostas semiparamétricas para estimação do parâmetro de memória em processos ARFIMA contaminados (Fajardo, Reisen & Cribari-Neto (2009)). O desempenho dos estimadores é analisado através de simulações de Monte Carlo e os resultados evidenciam a superioridade dos estimadores tipo M para aplicações práticas. Os estimadores robustos do parâmetro de memória podem ser considerados alternativas razoáveis para estimação em processos ARFIMA contaminados com dados atípicos.

ABSTRACT

This document focus on the study of methodologies used for estimating the long memory parameter in ARFIMA processes contaminated by atypical data. The suggested methodologies are based on robust estimators for the spectrum of the process, which can be reasonable alternatives for the estimation of the fractional difference parameter in contaminated processes. The robust estimators of the spectrum are classified in two classes: M -estimators and L -estimators. In the former class, the robust estimator of the spectrum is a function of the square sum of the process' Fourier coefficients. In the second class, the estimator demands the computing of a robust estimator for the autocovariance function, based on L -estimators (see, Ma & Genton (2000)). The computing of robust estimators for the spectrum of the process allows the development of semiparametric alternatives for estimating the memory parameter in contaminated ARFIMA processes (Fajardo et al. (2009)). The performance of the estimators is analyzed through Monte Carlo simulations and the results evidence the superiority of the M -estimators in practical applications. The robust estimators of the memory parameter can be considered as reasonable alternatives for estimating in ARFIMA processes contaminated by atypical data.

Capítulo 1

Introdução

O processo ARFIMA apresenta-se como uma extensão dos processos Autorregressivos e de Médias Móveis Integrados (ARIMA), onde o parâmetro de integração assume valores fracionários (ver, e.g. Granger & Joyeux (1980) e Hosking (1981)). Séries com representação $ARFIMA(p, d, q)$, com as raízes dos polinômios autorregressivo e de médias móveis fora do círculo unitário, são estacionárias mas não invertíveis quando $d \leq -0.5$. Para $|d| < 0.5$ as séries são estacionárias e invertíveis; e quando $d \geq 0.5$ o processo ARFIMA é não-estacionário. Hosking (1981) mostrou que para valores $d > 0$, séries com representação ARFIMA apresentam a propriedade de memória longa, que usualmente refere-se às correlações estatisticamente significativas entre observações distantes; ou equivalentemente, à singularidade na frequência zero da função de densidade espectral.

Aplicações empíricas dos processos ARFIMA têm sido encontradas em diversas áreas, por exemplo economia e finanças (ver, e.g. Bhardwaj & Swanson (2006), Reisen, Cribari-Neto & Jensen (2003) e Teyssière & Kirman (2007)), climatologia (ver, e.g. Baillie & Chung (2002)) entre outras. Para uma revisão recente de processos com propriedade de memória longa ver Doukhan, Oppenheim & Taqqu (2003) e Palma (2007).

Métodos propostos para estimação dos parâmetros do modelo ARFIMA são classificados em paramétricos e semi-paramétricos. Os métodos paramétricos consistem da estimação simultânea dos parâmetros do modelo, em geral por máxima verossimilhança; ver, e.g., Beran (1995), Dahlhaus (1989), Fox & Taqqu (1986), Hauser (1999) e Sowell (1992). No procedimento semi-paramétrico, a estimação dos parâmetros do modelo é realizada em dois passos: primeiro estima-se d e posteriormente estimam-se os parâmetros autorregressivos e de médias móveis. O estimador mais conhecido dentro dessa classe foi proposto por Geweke & Porter-Hudak (1983), obtido através de um modelo de regressão linear do logaritmo da função periodograma e denotado por LS . Variantes desse estimador foram desenvolvidas por Lobato & Robinson (1996), Reisen (1994), Robinson(1995*a*, 1995*b*), Velasco (2000), entre outros.

Hurvich, Deo & Brodsky (1998) e Robinson (1995*b*) provaram a consistência e normalidade assintótica do estimador LS em processos estacionários. Na região de não-estacionariedade, o estimador apresenta propriedades estatísticas diferentes para diferentes valores do d . Por exemplo, para $d \in [\frac{1}{2}, 1]$ o estimador é consistente; apresenta distribuição assintoticamente normal para $d \in [\frac{1}{2}, \frac{3}{4}]$. Para a região $d \in (\frac{3}{4}, 1]$ o estimador não apresenta distribuição normal (cf. Phillips(1999, 2007), Kim & Phillips (2006) e Velasco (1999*a*, 1999*b*)).

Estudos de simulação comparando diferentes técnicas de estimação de parâmetros em modelos ARFIMA, para diversos cenários, podem ser encontrados, por exemplo, em Bisaglia & Guégan (1998), Reisen, Abraham & Toscano(2000, 2002), Reisen, Abraham & Lopes (2001), Reisen, Rodrigues & Palma (2006), Smith, Taylor & Yadav (1997) entre

outros.

Na análise de séries temporais é comum encontrar observações influenciadas por eventos externos que podem facilmente afetar os procedimentos convencionais de análise, e podem enviesar significativamente as estimativas dos parâmetros do modelo. O efeito dessas observações, conhecidas como dados atípicos ou *outliers*, é no entanto, muitas vezes omitido pela falta do conhecimento de métodos que podem ser usados para detectá-los e para acomodá-los ao processo subjacente à série. Em aplicações práticas, a consideração dos dados atípicos nos procedimentos de modelagem tem como objetivo melhorar a eficiência do modelo ajustado e proporcionar uma maior confiabilidade em seus resultados (ver, e.g. Reisen, Fajardo & Teixeira (2008)).

Fox (1972) introduziu o conceito de *outliers* no contexto de séries temporais tendo considerado dois tipos de observações atípicas, a saber: aditivo (“Additive Outlier”) e inovador (“Innovational Outlier”). Chang, Tiao & Chen (1988), Chen & Liu (1993b, 1993a) e Tsay (1986) consideraram as alterações na estrutura da série, nomeadamente, alteração de nível permanente (“Level Shift”) e alterações temporárias (“Temporary Change”). Os mesmos autores, adotando a formulação de Fox (1972), consideraram esses dados como casos particulares do modelo geral de intervenção de Box & Tiao (1975).

Os efeitos dos *outliers* têm sido estudados por varios autores, por exemplo, Ledolter (1989) mostrou que intervalos de previsão são sensíveis a presença de *outliers* aditivos nos dados, mas previsões pontuais não são significativamente afetadas por esse tipo de dados, a não ser quando os *outliers* encontram-se próximos da origem da previsão; Chang et al. (1988) e Chen & Liu(1993a, 1993b) mostraram que a presença de *outliers* nos dados provoca viés

nas estimativas dos parâmetros do modelo ARMA; Chan(1992, 1995), Deutsch, Richards & Swain (1990), Fajardo et al. (2009) e Haldrup & Nielsen (2007) derivaram resultados sobre o viés ocasionado na função de autocorrelação amostral pela presença de observações atípicas. Os autores concluíram que o efeito das observações atípicas varia dependendo da posição dessas observações ao longo da série temporal. Além disso, os mesmos mostraram que a presença de *outliers* de tipo aditivo causa mudanças nas propriedades estatísticas da covariância amostral, e conseqüentemente no periodograma. O uso do periodograma nos procedimentos de estimação afeta significativamente as estimativas dos parâmetros quando existe a suspeita de contaminação na série temporal.

O estudo de modelos de memória longa na presença de *outliers* tem sido, recentemente, um assunto de muito interesse para pesquisadores da área. Evidências empíricas mostram que as estimativas do parâmetro de memória são significativamente alteradas pela presença de *outliers* do tipo aditivo na série temporal. Haldrup & Nielsen (2007) mostraram, através de simulações de Monte Carlo, que a presença de erros de medição, *outliers* e mudanças estruturais em séries temporais com tamanhos de amostras relativamente pequenos afetam seriamente as estimativas obtidas para o parâmetro de memória causando um viés nas estimativas do parâmetro de integração fracionária. Os autores concluíram que os estimadores semi-paramétricos obtidos por regressão apresentam viés relativamente menor quando o *bandwidth*, que corresponde ao número de frequências utilizadas para o cálculo das estimativas, é reduzido. Fajardo et al. (2009) mostraram que o viés pode ser explicado por uma translação da função de densidade espectral do processo observado.

Para diminuir o efeito causado pelas observações atípicas na estimação do espectro,

Martin & Thomson (1982) sugerem um algoritmo baseado na aplicação de um filtro de “limpeza” (ou cleaner-filter) no conjunto de observações. O filtro permite reduzir o impacto de cada possível outlier detectado ao longo do conjunto de observações, substituindo-o por uma previsão robusta do valor atípico. O desenvolvimento do filtro envolve a estimação robusta dos coeficientes de aproximações autorregressivas do processo observado para obter as previsões dos dados. Variantes para o filtro de “limpeza” são propostas por Tatum & Hurvich (1993) e Spangl & Dutter (2005). A aplicação dos filtros transforma a sequência de valores em uma nova série temporal sem contaminação, porém o espectro da série transformada pode diferir do espectro da série original, mesmo quando essa última não apresenta contaminação.

Fajardo et al. (2009) sugerem um periodograma robusto alisado para estimar o espectro de processos estacionários contaminados com outliers de tipo aditivo. O periodograma alisado baseia-se em um estimador robusto para a função de autocovariância proposto por Ma & Genton (2000). Para reduzir a contribuição dos termos finais da autocovariância amostral robusta, os autores sugerem o uso da *janela* truncada para o alisamento do periodograma, causando uma diminuição na variância do estimador e tornando-o estatisticamente consistente. Resultados empíricos mostram o desempenho do estimador na estimação dos parâmetros de processos estacionários na presença de dados atípicos. O estimador alisado pode ser considerado uma alternativa robusta razoável para estimação do espectro. O estudo empírico, conduzido por Fajardo (2011), mostra que a utilização de *janelas* de alisamento alternativas permite melhorar as propriedades estatísticas do periodograma robusto.

Em processos ARFIMA contaminados por dados atípicos, várias propostas para

estimação robusta do parâmetro d podem ser encontradas na literatura. Por exemplo, Agostinelli & Bisaglia (2003) sugeriram um método paramétrico baseado em verossimilhança ponderada como uma modificação do estimador proposto por Beran (1994). Fajardo et al. (2009) propuseram uma metodologia robusta para estimar os parâmetros do modelo ARFIMA na presença de *outliers* do tipo aditivo, baseada em um estimador *plugin* do estimador LS , onde a função periodograma é substituída por um periodograma robusto alisado com a *janela* truncada. Reisen & Fajardo (2008) propuseram uma extensão da metodologia robusta na estimação de parâmetros fracionários sazonais em processos com propriedades de memória longa.

Na estimação de processos ARFIMA não-estacionários na presença de observações atípicas, as propriedades estatísticas do estimador robusto, sugerido por Fajardo et al. (2009), são comprometidas; e a estimação do parâmetro de memória através desse estimador não é recomendável para aplicações práticas. No caso não-estacionário, o estimador robusto é invariante para séries com primeira diferença. Lopes, Olbermann & Reisen (2004) através de resultados empíricos mostram que o estimador sugerido por Geweke & Porter-Hudak (1983) mantém a mesma propriedade para séries sem contaminação. O alto custo computacional do estimador robusto para o cálculo das estimativas do parâmetro d , assim como a dificuldade para mostrar as propriedades assintóticas do mesmo, motivam o desenvolvimento de metodologias alternativas para o melhoramento das propriedades do estimador do parâmetro de memória em processos ARFIMA contaminados.

Visando propor uma técnica que ofereça maior eficiência nas aplicações empíricas, neste trabalho sugere-se um estimador robusto do parâmetro de memória de processos

ARFIMA estacionários e não-estacionários. O estimador baseia-se em uma variante do estimador LS , onde o estimador do espectro resulta da soma de quadrados dos M -estimadores dos coeficientes de Fourier do processo. Resultados empíricos através de simulações de Monte Carlo sugerem que as propriedades estatísticas do novo estimador se mostram equivalentes para tamanhos de amostra suficientemente grandes em séries não-contaminadas. O novo estimador do parâmetro de memória é uma alternativa eficaz quando a série temporal é contaminada por dados atípicos de tipo aditivo. Destaca-se o desempenho do estimador robusto do parâmetro d mesmo quando as séries não são contaminadas, sugerindo que pode ser considerado uma alternativa razoável para estimação do parâmetro de diferenciação fracionária em séries sem contaminação.

Este documento encontra-se organizado da seguinte forma: No primeiro artigo, intitulado “*Estimadores robustos para o espectro de processos estacionários*”, sugere-se um periodograma robusto para a estimação do espectro de processos estacionários na presença de dados atípicos de tipo aditivo. O estimador é chamado de periodograma tipo M e baseia-se na soma de quadrados dos M -estimadores dos coeficientes de Fourier da representação espectral de processos estacionários. As propriedades estatísticas do estimador indicam que o mesmo é assintoticamente não-viesado. Estudos empíricos, através de simulações de Monte Carlo, comparam o desempenho do estimador robusto para diferentes funções de perda: Laplace (ou L_1), *bisquare*, Hampel e Huber. As funções *bisquare* e Hampel destacam-se por apresentar melhor desempenho em amostra finita, quanto às medidas de erro de precisão e eficiência relativa.

Finalmente, no segundo artigo intitulado “*Estimadores robustos para o parâmetro de*

diferenciação fracionária em processos ARFIMA(p, d, q)” propõe-se um estimador robusto para o parâmetro de memória em processos ARFIMA estacionários e não-estacionários. Aplicações empíricas mostram que estimadores tipo M e L apresentam um desempenho razoável em processos estacionários com contaminação. No caso não-estacionário, na presença de dados atípicos o estimador baseado em periodogramas tipo L devem ser evitados. Os resultados sugerem que para cenários contaminados estimadores tipo M resultam uma opção razoável na estimação dos parâmetros de modelos ARFIMA.

Referências Bibliográficas

- Agostinelli, C. & Bisaglia, L. (2003), Robust estimation of ARFIMA processes, Technical report, Università Ca' Foscari di Venezia.
- Baillie, R. T. & Chung, S.-K. (2002), 'Modeling and forecasting from trend-stationary long memory models with applications to climatology', *International Journal of Forecasting* **18**, 215–226.
- Beran, J. (1994), 'On a class of M-estimators for gaussian long-memory models', *Biometrika* **81**, 755–766.
- Beran, J. (1995), 'Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models', *Journal of the Royal Statistical Society* **57**(B), 659–672.
- Bhardwaj, G. & Swanson, N. (2006), 'An empirical investigation of the usefulness of ARFIMA models for predicting macroeconomic and financial time series', *Journal of Econometrics* **131**, 539–578.
- Bisaglia, L. & Guégan, D. (1998), 'A comparison of techniques of estimation in long-memory process', *Computational Statistics & Data Analysis* **27**, 61–81.

- Box, G. E. P. & Tiao, G. C. (1975), 'Intervention analysis with applications to economic and environmental problems', *Journal of the American Statistical Association* **70**(349), 70–79.
- Chan, W. (1992), 'A note on time series model specification in the presence outliers', *Journal of Applied Statistics* **19**, 117–124.
- Chan, W. (1995), 'Outliers and financial time series modelling: a cautionary note', *Mathematics and Computers in Simulation* **39**, 425–430.
- Chang, I., Tiao, G. C. & Chen, C. (1988), 'Estimation of time series parameters in presence of outliers', *Technometrics* **30**, 1936–204.
- Chen, C. & Liu, L. (1993a), 'Forecasting time series with outliers', *Journal of Forecasting* **12**, 13–35.
- Chen, C. & Liu, L. (1993b), 'Joint estimation of model parameters and outlier effects in time series', *Journal of the American Statistical Association* **88**, 284–297.
- Dahlhaus, R. (1989), 'Efficient parameter estimation for self-similar processes', *The Annals of Statistics* **17**, 1749–1766.
- Deutsch, S. J., Richards, J. E. & Swain, J. J. (1990), 'Effects of a single outlier on ARMA identification', *Communications in Statistics: Theory and Methods* **19**, 2207–2227.
- Doukhan, P., Oppenheim, G. & Taqqu, M. (2003), *Theory and Applications of Long-Range Dependence*, Birkhäuser.

- Fajardo, F. (2011), ‘On some alternatives for robust estimation of the spectrum in stationary processes’, *Brazilian Review of Econometrics* (available in <http://bibliotecadigital.fgv.br/ojs/index.php/bre/article/view/2767>) .
- Fajardo, F., Reisen, V. A. & Cribari-Neto, F. (2009), ‘Robust estimation in long-memory processes under additive outliers’, *Journal of Statistical Planning and Inference* **139**, 2511–2525.
- Fox, A. J. (1972), ‘Outliers in time series’, *Journal of the Royal Statistical Society* **34**(B), 350–363.
- Fox, R. & Taqqu, M. S. (1986), ‘Large-sample properties of parameters estimates for strongly dependent stationary gaussian time series’, *The Annals of Statistics* **14**, 517–532.
- Geweke, J. & Porter-Hudak, S. (1983), ‘The estimation and application of long memory time series model’, *Journal of Time Series Analysis* **4**, 221–238.
- Granger, C. W. J. & Joyeux, R. (1980), ‘An introduction to long-memory time series models and fractional differencing’, *Journal of Time Series Analysis* **1**, 15–30.
- Haldrup, N. & Nielsen, M. O. (2007), ‘Estimation of fractional integration in the presence of data noise’, *Computational Statistics & Data Analysis* **51**, 3100–3114.
- Hauser, M. (1999), ‘Maximum likelihood estimators for ARMA and ARFIMA models: a monte carlo study’, *Journal of Statistical Planning and Inference* **80**, 229–255.
- Hosking, J. R. (1981), ‘Fractional differencing’, *Biometrika* **68**, 165–176.

- Hurvich, C. M., Deo, R. & Brodsky, J. (1998), ‘The mean square error of Geweke and Porter-Hudak’s estimator of the memory parameter of a long-memory time series’, *Journal of Time Series Analysis* **19**(1), 19–46.
- Kim, C. S. & Phillips, P. (2006), Log periodogram regression: the nonstationary case, Technical report, Cowles Foundation Discussion Paper, Yale University.
- Ledolter, J. (1989), ‘The effect of additive outliers on the forecast from ARMA models’, *International Journal of Forecasting* **5**, 231–240.
- Lobato, I. & Robinson, P. M. (1996), ‘Averaged periodogram estimation of long memory’, *Journal of Econometrics* **73**, 303–324.
- Lopes, S. R. C., Olbermann, B. P. & Reisen, V. A. (2004), ‘Comparison of estimation methods in non-stationary ARFIMA process’, *Journal of Statistical Computation and Simulation* **74**(5), 339–347.
- Ma, Y. & Genton, M. (2000), ‘Highly robust estimation of the autocovariance function’, *Journal of Time Series Analysis* **21**, 663–684.
- Martin, R. D. & Thomson, D. (1982), ‘Robust-resistant spectrum estimation’, *Proceedings of the IEEE* **70**(9), 1097–1115.
- Palma, W. (2007), *Long-Memory Time Series: Theory and Methods*, Wiley-Interscience.
- Phillips, P. (1999), Discrete fourier transforms of fractional processes, Technical report, Cowles Foundation Discussion Paper, Yale University.

- Phillips, P. (2007), ‘Unit root log periodogram regression’, *Journal of econometrics* **138**, 104–124.
- Reisen, V. A. (1994), ‘Estimation of the fractional difference parameter in the ARIMA(p, d, q) model using the smoothed periodogram’, *Journal of Time Series Analysis* **15**, 335–350.
- Reisen, V. A., Abraham, B. & Lopes, S. (2001), ‘Estimation of parameters in ARFIMA processes: A simulation study’, *Communications in Statistics: Simulation and Computation* **30**, 787–803.
- Reisen, V. A., Abraham, B. & Toscano, E. (2000), ‘Parametric and semiparametric estimations of stationary univariate ARFIMA models’, *Brazilian Journal of Probability and Statistics* **14**, 185–206.
- Reisen, V. A., Abraham, B. & Toscano, E. (2002), ‘Effect of parameter estimation on estimating the forecast error variace in an ARFIMA processes: a simulation study and an example’, *Statistical Methods* **4**, 21–37.
- Reisen, V. A., Cribari-Neto, F. & Jensen, M. (2003), ‘Long memory inflationary dynamics: The case of Brazil’, *Studies in Nonlinear Dynamics & Econometrics* **7**, 1–16.
- Reisen, V. A. & Fajardo, F. (2008), Robust estimation in seasonal long-memory processes with outliers, in ‘Annals of Latin American Meeting of the Econometric Society’, <http://www.webmeets.com/files/papers/LACEA-LAMES/2008/707/SeasonalOutlierV2.pdf>.

- Reisen, V. A., Fajardo, F. & Teixeira, E. (2008), ‘Modelagem de séries temporais sazonais na presença de outliers: Estudo de caso da vazão máxima mensal do rio Jucu, ES, Brasil’, *Revista Brasileira de Recursos Hídricos* **13**, 45–53.
- Reisen, V. A., Rodrigues, A. & Palma, W. (2006), ‘Estimating seasonal long-memory processes: A Monte Carlo study’, *Journal of Statistical Computation and Simulation* **76**(4), 305–316.
- Robinson, P. M. (1995a), ‘Gaussian semiparametric estimation of long range dependence’, *The Annals of Statistics* **23**, 1630–1661.
- Robinson, P. M. (1995b), ‘Log-periodogram regression of time series with long range dependence’, *The Annals of Statistics* **23**, 1048–1072.
- Smith, J., Taylor, N. & Yadav, S. (1997), ‘Comparing the bias and misspecification in ARFIMA models’, *Journal of Time Series Analysis* **18**, 507–527.
- Sowell, F. B. (1992), ‘Maximum likelihood estimation of stationary univariate fractionally integrated time series models’, *Journal of Econometrics* **53**, 165–188.
- Spangl, B. & Dutter, R. (2005), ‘On robust estimation of power spectra’, *Austrian Journal of Statistics* **34**(2), 199–210.
- Tatum, L. & Hurvich, C. (1993), ‘High breakdown methods of time series analysis’, *Journal of the Royal Statistical Society B* **55**, 881–896.
- Teyssière, G. & Kirman, A. (2007), *Long Memory in Economics*, Springer.

Tsay, R. S. (1986), 'Time series model specification in the presence of outliers', *Journal of the American Statistical Association* **81**, 132–141.

Velasco, C. (1999a), 'Gaussian semiparametric estimation of non-stationary time series', *Journal of Time Series Analysis* **20**(1), 87–127.

Velasco, C. (1999b), 'Non-stationary log-periodogram regression', *Journal of Econometrics* **91**, 325–371.

Velasco, C. (2000), 'Non-gaussian log-periodogram regression', *Econometric Theory* **16**, 44–79.

ESTIMADORES ROBUSTOS PARA O ESPECTRO DE PROCESSOS ESTACIONÁRIOS

FABIO A. FAJARDO - VALDÉRIO A. REISEN

Departamento de Estatística, UFMG, Belo Horizonte/MG, Brasil

Departamento de Estatística, UFES, Vitória/ES, Brasil

RESUMO. Neste trabalho estuda-se uma classe de estimadores robustos para o espectro de processos estacionários contaminados por outliers de tipo aditivo. A classe de estimadores é chamada de tipo M e baseia-se na soma de quadrados de estimadores robustos dos coeficientes de Fourier da representação espectral de processos estacionários. Os estimadores dos coeficientes são obtidos através da utilização de M -estimadores com diferentes funções de perda. Resultados teóricos mostram as propriedades assintóticas do periodograma tipo M. Estudos de Monte Carlo comparam empiricamente o desempenho dos estimadores tipo M e tipo L, sugerido por Fajardo et al. (2009), na estimação da densidade espectral de processos ARFIMA(p, d, q).

Palavras-chave. Espectro, periodograma, robustez, outliers.

1. INTRODUÇÃO

A análise de séries temporais no domínio da frequência baseia-se no estudo da função de densidade espectral (ou espectro). A densidade espectral de um processo estacionário pode ser interpretada como a proporção de variância atribuída às oscilações da série temporal em uma frequência dada. Na prática, um estimador para a densidade espectral de $\{y_t; t = 1, 2, \dots, n\}$ é o periodograma, definido como

$$I(\lambda_j) = \frac{1}{n} \left| \sum_{t=1}^n y_t \exp\{-i\lambda_j t\} \right|^2, \quad (1)$$

onde $\exp\{-i\lambda_j t\} = \cos(\lambda_j t) - i \sin(\lambda_j t)$ e $\lambda_j = \frac{2\pi j}{n}$ são chamadas de frequências fundamentais (ou de Fourier) com $0 < j < \lfloor \frac{n}{2} \rfloor$. O símbolo $\lfloor \cdot \rfloor$ denota a função parte inteira. Aplicações empíricas do periodograma podem ser encontradas nas diferentes áreas do conhecimento, frequentemente para a identificação de periodicidades ocultas na série temporal ou na estimação dos parâmetros dos modelos ajustados, entre outras aplicações (ver, e.g., Brillinger 1981).

Na análise de séries temporais varios tipos de observações atípicas ou *outliers* (ver, Fox 1972) são comumente encontrados nas realizações dos processos observados e, dependendo da sua natureza, podem causar efeitos significativos no uso das metodologias tradicionais. Como mostrado por Fajardo et al. (2009) e Haldrup & Nielsen (2007) a presença de outliers de tipo aditivo causa mudanças nas propriedades estatísticas da covariância amostral e, conseqüentemente, no periodograma. Os autores mostram que o uso do periodograma nos procedimentos de estimação afeta significativamente as estimativas dos parâmetros quando existe a suspeita de contaminação na série temporal.

Para diminuir o efeito causado pelas observações atípicas na estimação do espectro, Martin & Thomson (1982) sugerem um algoritmo baseado na aplicação de um filtro de “limpeza” (ou cleaner-filter) no conjunto de observações. Esse filtro permite reduzir o impacto de cada possível outlier detectado ao longo do conjunto de observações, substituindo-o por uma previsão robusta do valor

Date: 29 de dezembro de 2011.

Email addresses: ffajardo@est.dout.ufmg.br.

atípico. O desenvolvimento do filtro envolve a estimação robusta dos coeficientes de aproximações autorregressivas do processo observado para obter as previsões dos dados. Variantes para o filtro de “limpeza” são propostas por Tatum & Hurvich (1993) e Spangl & Dutter (2005). A aplicação dos filtros transforma a sequência de valores em uma nova série temporal sem contaminação, porém o espectro da série transformada pode diferir do espectro da série original, mesmo quando essa última não apresenta contaminação.

Fajardo et al. (2009) sugerem um periodograma robusto alisado para o espectro de processos estacionários contaminados com outliers de tipo aditivo. O periodograma alisado baseia-se em um estimador robusto para a função de autocovariância proposto por Ma & Genton (2000). Para reduzir a contribuição dos termos finais da autocovariância amostral robusta, os autores sugerem o uso da *janela* truncada para o alisamento do periodograma, causando uma diminuição na variância do estimador e tornando-o estatisticamente consistente. Resultados empíricos mostram o desempenho do estimador na estimação dos parâmetros de processos estacionários na presença de dados atípicos. O estimador alisado pode ser considerado uma alternativa robusta razoável para estimação do espectro. O estudo empírico, conduzido por Fajardo (2011), mostra que a aplicação das *janelas* de Bartlett e Parzen permite melhorar as propriedades estatísticas do periodograma robusto. O desempenho do estimador é avaliado através de medidas de erro de precisão e eficiência relativa.

Neste trabalho sugere-se um periodograma robusto para a estimação do espectro de processos estacionários na presença de dados atípicos de tipo aditivo. O estimador é chamado de periodograma tipo M e baseia-se na soma de quadrados dos M-estimadores dos coeficientes de Fourier da representação espectral de processos estacionários. Os M-estimadores dos coeficientes são obtidos com diferentes funções de perda. As propriedades estatísticas do estimador indicam que o mesmo é assintoticamente não-viesado e com um ponto de quebra menor a 25%. Estudos empíricos, através de simulações de Monte Carlo, comparam o desempenho do estimador robusto para diferentes funções de perda: Laplace (ou L_1), *bisquare*, Hampel e Huber. As funções *bisquare* e Hampel destacam-se por apresentar melhor desempenho em amostra finita.

Este artigo encontra-se organizado da seguinte forma. A Seção 2 introduz os estimadores clássicos do espectro de processos estacionários. Na Seção 3 são apresentadas as definições e propriedades estatísticas dos estimadores robustos do espectro. Resultados numéricos e análise empírica são descritos na Seção 4. Finalmente, a Seção 5 apresenta algumas conclusões e comentários.

2. ESTIMADORES DA DENSIDADE ESPECTRAL

Seja $\{y_t; t \in \mathbb{Z}\}$ um processo estacionário com média 0 e função de densidade espectral dada por

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} R(h)e^{-ih\lambda}, \text{ com } \lambda \in [-\pi, \pi], \quad (2)$$

onde $R(h) = \mathbb{E}[y_t y_{t+h}]$ é a função de autocovariância de $\{y_t; t \in \mathbb{Z}\}$.

Seja $\{y_t; t = 0, 1, \dots, n\}$ uma realização do processo $\{y_t; t \in \mathbb{Z}\}$. Um estimador natural do espectro é definido por $\hat{f}(\lambda_j) = \frac{1}{2\pi} I(\lambda_j)$, tal que

$$I(\lambda_j) = \sum_{|h| < n} \hat{R}(h) e^{-ih\lambda_j}, \quad (3)$$

onde $\lambda_j = \frac{2\pi j}{n}$, com $j = 1, 2, \dots, \lfloor \frac{n}{2} \rfloor$ e $\hat{R}(h) = \frac{1}{n} \sum_{i=1}^{n-|h|} y_{i+|h|} y_i$ é um estimador para $R(h)$.

A função periodograma é assintoticamente não-viesada para a função $2\pi f(\lambda_j)$, com variância dada por $\text{var}[I(\lambda_j)] = (2\pi)^2 f^2(\lambda_j) + \mathcal{O}(n^{-1/2})$, com $0 < \lambda_j < \pi$ (para detalhes ver, e.g., Brockwell & Davis 2006). Versões alisadas do estimador definido na Eq. 3 são sugeridas para torná-lo estatisticamente

consistente. Um estimador alisado para a densidade espectral é dado por

$$\hat{f}_s(\lambda_j) = \frac{1}{2\pi} \sum_{|k| \leq r} w_n(k) I(\lambda_{j+k}), \quad (4)$$

onde r é chamado de *bandwidth* e satisfaz $0 < r < n$ com $r = o(n)$. A sequência $\{w_n(\cdot)\}$ satisfaz as seguintes condições: *i.*) $w_n(k) = w_n(-k)$ e $w_n(k) \geq 0$ para todo k , *ii.*) $\sum_{|k| \leq r} w_n(k) = 1$, *iii.*) $\sum_{|k| \leq r} w_n^2(k) \rightarrow 0$ quando $n \rightarrow \infty$. Em particular, quando $w_n(j) = \frac{1}{n} \sum_{|h| \leq r} \kappa\left(\frac{h}{r}\right) e^{-ih\lambda_j}$ obtem-se aproximadamente

$$\hat{f}_s(\lambda_j) = \frac{1}{2\pi} \sum_{|h| \leq r} \kappa\left(\frac{h}{r}\right) \hat{R}(h) e^{ih\lambda_j}, \quad (5)$$

onde a função $\kappa(\cdot)$ é chamada de *janela de defasagem* e satisfaz $\kappa(0) = 1$, $|\kappa(x)| \leq 1$ para todo x e $\kappa(x) = 0$ se $|x| > 1$ (para detalhes ver, e.g. Anderson (1971) e Priestley (1981)). Alguns exemplos de *janelas*:

- i. janela truncada* $\kappa(x) = 1$ se $|x| \leq r$, com $w_n(\theta) = D_n(\theta) := \frac{1}{2\pi} \frac{\text{sen}((r+1/2)\theta)}{\text{sen}(\theta/2)}$. A variância assintótica é dada por $\frac{2r}{n} f^2(\lambda)$. Para este tipo de *janela* não existe uma expressão fechada que permita avaliar o viés do estimador (Priestley 1981).
- ii. janela de Parzen*:

$$\kappa(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & \text{se } |x| \leq 0.5, \\ 2(1 - |x|)^3, & \text{se } 0.5 < |x| \leq 1.0, \\ 0, & \text{se } |x| > 1, \end{cases}$$

com $w_n(\theta) = \frac{6}{\pi r^3} \frac{\text{sen}^4(r\theta/4)}{\text{sen}^4(\theta/2)}$. A variância e o viés assintóticos são, respectivamente, dados por $0.5393 \frac{r}{n} f^2(\lambda)$ e $-\frac{6}{r^2} f^{(2)}(\lambda)$, onde $f^{(2)}(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} h^2 R(h) e^{-ih\lambda}$.

- iii. janela de Tuckey-Hamming* (TH): $\kappa(x) = 0.54 + 0.46 \cos(\pi x)$, $|x| \leq 1$, com $w_n(\theta) = 0.23D_n(\theta - \pi/r) + 0.54D_n(\theta) + 0.23D_n(\theta + \pi/r)$. A variância e o viés assintóticos são, respectivamente, dados por $0.7948 \frac{r}{n} f^2(\lambda)$ e $-\frac{0.23\pi^2}{r^2} f^{(2)}(\lambda)$.

- iv. janela de Bartlett*: $\kappa(x) = 1 - |x|$, $|x| \leq 1$, com $w_n = (\theta) \frac{1}{2\pi r} \frac{\text{sen}^2(r\theta/2)}{\text{sen}^2(\theta/2)}$. A variância e o viés assintóticos são, respectivamente, dados por $\frac{2r}{3n} f^2(\lambda)$ e $-\frac{1}{r} f^{(1)}(\lambda)$, onde $f^{(1)}(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} |h| R(h) e^{-ih\lambda}$.

Ao selecionar a *janela* alisadora deve-se levar em conta a escolha do valor do r . Na prática, sugere-se a utilização de critérios estatísticos adequados para seleção ótima de tal valor. Para detalhes ver, e.g., Beltrão & Bloomfield (1987), Hurvich & Beltrão (1990) e Robinson (1996) entre outros.

3. ESTIMADOR ROBUSTO PARA O ESPECTRO

A seguir é apresentada uma classe de estimadores robustos para a função de densidade espectral de um processo estacionário. O estimador é chamado de *periodograma tipo M* e calcula-se a partir da soma de quadrados dos M-estimadores dos coeficientes da representação espectral do processo. Para o cálculo dos M-estimadores sugere-se o uso de diferentes funções de perda (Huber 2004). As propriedades do estimador são apresentadas para processos estacionários com propriedades de memória curta e longa.

3.1. Periodograma tipo M. Seja $\{y_t; t = 1, 2, \dots, n\}$ um processo gaussiano estacionário com média zero e função de densidade espectral $f(\lambda)$, com $\lambda \in [-\pi, \pi]$. Um estimador robusto para a função $f(\lambda)$ é dado por $\tilde{f}(\lambda_j) = \frac{1}{2\pi} \tilde{I}(\lambda_j)$, onde

$$\tilde{I}(\lambda_j) = \frac{n}{4} \{T_{nj,0}^2 + T_{nj,1}^2\}, \quad (6)$$

$\lambda_j = \frac{2\pi j}{n}$ com $j = 1, 2, \dots, \lfloor \frac{n}{2} \rfloor$. Para cada j , o vetor $T_n(F_n; \lambda_j) := T_{nj} = (T_{nj,0}, T_{nj,1})'$ é um M-estimador para $\theta = (\theta_{0j}, \theta_{1j})'$ que satisfaz

$$\frac{1}{n} \sum_{t=1}^n \rho \left(\frac{y_t - \mathbf{x}_t' T_{nj}}{\hat{\sigma}} \right) = \min_{\theta \in \mathbb{R}^2} \frac{1}{n} \sum_{t=1}^n \rho \left(\frac{y_t - \mathbf{x}_t' \theta}{\hat{\sigma}} \right), \quad (7)$$

onde $\mathbf{x}_t = (\cos(\lambda_j t), \sin(\lambda_j t))'$, F_n representa a função de distribuição empírica e $\hat{\sigma}$ um estimador robusto para $\sigma = \sqrt{\text{var}(y_t - \mathbf{x}_t' \theta)}$. Estimativas robustas para σ podem ser obtidas através do desvio absoluto médio padronizado (ou MADN, abreviatura do inglês). Outras alternativas estatisticamente mais eficientes podem ser encontradas em Rousseeuw & Croux (1993). A função $\rho: \mathbb{R} \rightarrow \mathbb{R}$ é chamada de *função de perda* e pode ser qualquer função convexa que satisfaz certas condições para diferenciabilidade e permita obter estimativas robustas para o vetor θ . Em geral, a convexidade na função ρ não é uma condição necessária para desenvolver a teoria de M-estimadores (ver, e.g. Haberman 1989). No entanto, Huber (1964) mostra que a convexidade da função ρ permite provar a consistência e normalidade assintótica do estimador T_{nj} sob condições mínimas, assim como unicidade e existência da solução da Eq. 7.

A definição de M-estimador na Eq. 7 deve-se a Huber (1964) e diferentes aplicações deste estimador têm sido desenvolvidas na literatura; várias referências podem ser encontradas nas publicações de Hampel et al. (1986), Huber (2004), Maronna et al. (2006) e Bai et al. (1997) entre outros.

3.1.1. Propriedades do estimador. O estimador definido na Eq. 6 tem sido definido para um conjunto discreto de frequências $\lambda_j = \frac{2\pi j}{n}$. Para estudar as propriedades assintóticas do estimador é conveniente definir um $\lambda \in (0, \pi]$ tal que $\frac{\pi(2j-1)}{n} < \lambda \leq \frac{\pi(2j+1)}{n}$ para $j = 1, 2, \dots$ (ver, e.g. Fuller (1996, p. 356)). Nesse caso $\tilde{I}(\lambda)$, com $\lambda \in (0, \pi]$, é uma função degrau que toma valores $\tilde{I}(\lambda_j)$ no intervalo $\left(\frac{\pi(2j-1)}{n}; \frac{\pi(2j+1)}{n} \right)$. As condições, sob convexidade da função ρ , que garantem a consistência e normalidade assintótica do estimador $T_n = T_n(F_n, \lambda)$ são apresentadas a seguir.

Suposição 1. *Existe uma função $\varrho(\cdot)$ subgradiente de $\rho(\cdot)$ tal que $\mathbb{E}[\varrho(e_1)] = 0$, $0 < \mathbb{E}[\varrho^2(e_1)] < \infty$ e $\mathbb{E}[|\rho(e_1+c) - \rho(e_1) - c\varrho(e_1)| \wedge |\rho(e_1+c) - \rho(e_1) - c\varrho(e_1)|^2] = o(c^2)$, onde $e_1 = y_1 - \mathbf{x}_1' \theta$, $a \wedge b := \min\{a, b\}$. Uma função $\varrho(u)$ se diz subgradiente de $\rho(u)$ se $\rho(z) \geq \rho(u) + (z - u)' \varrho(u)$ para todo $z \in \mathbb{R}^2$ (ver, e.g. Rockafellar (1981)).*

Suposição 2. *Suponha $\mathbb{E}[\rho(e_1 + c) - \rho(e_1)] > 0$ para cada $c \neq 0$ e suponha que existe $k > 0$ tal que $\mathbb{E}[\rho(e_1 + c) - \rho(e_1)] = kc^2 + o(c^2)$ quando $c \rightarrow 0$.*

Suposição 3. *Suponha $\mathbb{E}[\rho(e_1 - c)]$ uma função contínua em c e*

$$\lim_{\xi \rightarrow 0} \mathbb{E} \left[\sup_{|s| \leq \xi} |\rho(e_1 - \mathbf{x}_1'(\theta - s)) - \rho(e_1 - \mathbf{x}_1' \theta)| \right] = 0.$$

As Suposições 1 e 2 são satisfeitas para qualquer função convexa, duplamente diferenciável com derivadas contínuas, tal que $0 < \mathbb{E}[\varrho^2(e_1)] < \infty$. As condições acima definidas garantem a existência de uma única solução para a Eq. 7 e permitem mostrar que a estatística T_{nj} também é solução do sistema de equações dado por

$$\sum_{t=1}^n \varrho \left(\frac{y_t - \mathbf{x}_t' \theta}{\hat{\sigma}} \right) \mathbf{x}_t = 0, \quad (8)$$

ou seja, o M-estimador T_{nj} pode ser encontrado como solução das Eqs. 7 ou 8 de forma equivalente. A Suposição 3 tem um papel puramente teórico e requer-se para provar a consistência estatística do estimador T_{nj} (ver, Arcones 2001).

As condições apresentadas nas Suposições 1-3 são sugeridas por Arcones (1998, 2001) e Bai et al. (1992) para abrangência de uma classe ampla de funções de perda sob requisitos mínimos.

Huber (2004), Serfling (1980) e Yohai & Maronna (1979) entre outros apresentam resultados semelhantes para consistência e normalidade assintótica sob condições mais rigorosas e restritas, que não abrangem casos especiais para a função de perda, por exemplo $\rho(x) = |x|^p$ com $p \geq 1$, onde $\varrho(x) = p|x|^{p-1} \text{sgn}(x)$ e $\text{sgn}(x) = \mathbf{1}(x > 0) - \mathbf{1}(x < 0)$, onde $\mathbf{1}(\cdot)$ é a função indicadora.

Os resultados da Proposição 1 permitem obter a distribuição assintótica do periodograma robusto, definido na Eq. 6, e apresentada no Corolário 1.

Proposição 1. *Seja ρ definida como na Eq. 7. Sob as Suposições 1 – 3, tem-se que*

- i. $T_n \xrightarrow{q.c.} \theta$, quando $n \rightarrow \infty$.
- ii. O estimador T_n é assintoticamente normal, i.e.

$$n^{1/2} (T_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tilde{\sigma}^2 \mathcal{I}_2) \text{ quando } n \rightarrow \infty,$$

onde $\tilde{\sigma}^2 = \frac{2}{k^2} \mathbb{E} [\varrho^2(e_1)]$ e \mathcal{I}_2 representa a matriz identidade 2×2 . Os símbolos “ $\xrightarrow{q.c.}$ ” e “ $\xrightarrow{\mathcal{D}}$ ” denotam convergências quase certa e em distribuição, respectivamente.

O valor de k é obtido a partir da relação dada na Suposição 2 e depende da forma da função de perda ρ . O item ii. da Proposição 1 sugere que as transformadas seno e coseno são assintoticamente independentes, para $\lambda \in (0, \pi)$, e a variância assintótica do estimador robusto depende do subgradiente da função de perda e, conseqüentemente, do valor da frequência λ_j . Os resultados apresentados na Proposição 1 são amplamente conhecidos na literatura e têm sido considerados por varios autores no contexto de regressão linear, ver, e.g. Arcones(1998, 2001), Bai et al.(1992, 1997) entre outros.

Corolário 1. *Sob as suposições dadas na Proposição 1, $\frac{n}{4} \{(T_{n0} - \theta_0)^2 + (T_{n1} + \theta_1)^2\}$ converge em distribuição para uma variável aleatória com distribuição $\frac{1}{4} \tilde{\sigma}^2 \chi_2^2$, onde χ_2^2 representa uma variável aleatória com distribuição qui-quadrado com 2 graus de liberdade.*

O resultado apresentado no corolário é consequência imediata da Proposição 1. Quando $\theta_0 = \theta_1 = 0$, o valor esperado assintótico do periodograma robusto é dado por $\mathbb{E} [\tilde{I}(\lambda)] = \frac{\mathbb{E} \varrho^2(e_1)}{k^2}$ e a variância assintótica dada por $\text{var} [\tilde{I}(\lambda)] = \left[\frac{\mathbb{E} \varrho^2(e_1)}{k^2} \right]^2$. No caso clássico, definido na Eq. 1 e denotado por $I_{LS}(\lambda)$, pode ser obtido quando o estimador T_n satisfaz a Eq. 7 com função de perda dada por $\rho(x) = \frac{1}{2}x^2$, ou equivalentemente T_n é solução da Eq. 8 com $\varrho(x) = x$. A distribuição assintótica do periodograma clássico é qui-quadrado com dois graus de liberdade e variância assintótica dada por $\text{var} [I_{LS}(\lambda)] = \sigma^4$, onde $\sigma^4 = (2\pi)^2 f^2(\lambda)$ com $\lambda \in (0, \pi)$ (para detalhes ver, e.g. Brockwell & Davis 2006).

O periodograma L_p obtem-se a partir da Eq. 6 quando o estimador T_n é solução da Eq. 8 com $\varrho(x) = p|x|^{p-1} \text{sgn}(x)$. Em particular, quando $p = 1$, o periodograma L_p é chamado de *periodograma de Laplace*, denotado por $I_{L_1}(\lambda)$. A distribuição assintótica de $I_{L_1}(\lambda)$ é dada por

$$I_{L_1}(\lambda) \xrightarrow{\mathcal{D}} \frac{1}{2} \frac{\sigma^2}{[2F'_e(0)]^2} \chi_2^2,$$

onde $\mathbb{E} [\varrho^2(e_1)] = \sigma^2$, $k = 2F'_e(0)$ e $F_e(\cdot)$ a função de distribuição de e_1 , tal que $F'_e(0) > 0$ e $F_e(0) = 0.5$. Li (2008) obtem resultados equivalentes para a distribuição do periodograma L_1 , quando $\sigma^2 = 1$. Aplicações do periodograma L_1 no contexto de processos com distribuição de caudas pesadas podem ser encontradas nos trabalhos de Katkovnik (1998) e Li (2008) entre outros. Para $p > 1$, a distribuição assintótica do periodograma L_p é dada por

$$I_{LP}(\lambda) \xrightarrow{\mathcal{D}} \frac{1}{2} \frac{\mathbb{E} [\varrho^2(e_1)]}{k^2} \chi_2^2,$$

onde $\mathbb{E} [\varrho^2(e_1)] = \int p^2 |e_1|^{2(p-1)} dF(e_1)$ e $k = \int p(p-1) |e_1|^{p-2} dF(e_1)$.

A definição do periodograma robusto na Eq. 6 permite sugerir alternativas para estimação do espectro de processos estacionários na presença de outliers. Nesta tese é analisado o periodograma robusto definido a partir do estimador T_n , que minimiza a Eq. 7, quando a função ρ é dada pelas funções *bisquare*, Huber e Hampel (para detalhes ver, e.g. Hampel et al. (1986), Huber (2004) e Maronna et al. (2006)).

Para tamanhos de amostra suficientemente grandes, o primeiro momento do estimador tende a 2π vezes a densidade espectral, sugerindo que o estimador robusto definido como $\tilde{f}(\lambda) = \frac{1}{2\pi}\tilde{I}(\lambda)$ com $\lambda \in (0, \pi)$, é um estimador assintoticamente não-viesado do espectro de um processo estacionário. O resultado é consequência imediata da relação $\frac{\mathbb{E}\varrho^2(\epsilon_1)}{k^2} \rightarrow 2\pi f(\lambda)$ quando $n \rightarrow \infty$, i.e. o resultado sugere que as propriedades assintóticas do estimador clássico podem ser estendidas de forma natural para o estimador robusto.

A aplicação de transformações não-lineares no periodograma é comumente encontrada nos procedimentos de estimação dos parâmetros de modelos ajustados (ver, e.g. Chen & Hannan (1980), Taniguchi (1979) e Geweke & Porter-Hudak (1983) entre outros). Uma transformação amplamente utilizada é o logaritmo natural. A seguir são apresentados o valor esperado e a variância do logaritmo natural do periodograma robusto.

Corolário 2. *Seja $\tilde{I}(\lambda)$ o estimador robusto, definido na Eq. 6, com $\lambda \in (0, \pi)$. O valor esperado e a variância do logaritmo do periodograma robusto padronizado, i.e. $\log \frac{2\tilde{I}(\lambda)}{\tilde{\sigma}^2}$, são dados por*

$$\mathbb{E} \left[\log \frac{2\tilde{I}(\lambda)}{\tilde{\sigma}^2} \right] = \psi(1) + \frac{1}{2} \log 2; \quad \text{var} \left[\log \frac{2\tilde{I}(\lambda)}{\tilde{\sigma}^2} \right] = \frac{\pi^2}{6},$$

quando $n \rightarrow \infty$. O valor $\psi(1) = 0.577216\dots$ representa a constante Euler.

O resultado acima é consequência imediata da expressão para o r -ésimo cumulante do logaritmo de uma variável aleatória qui-quadrado dado por $Cum_r(\log \chi_\nu^2) = \psi^{(r-1)}(0.5\nu) + \varepsilon_r \log 2$, onde $\varepsilon_1 = 0.5$ e $\varepsilon_r = 0$ para $r > 1$ (Johnson et al. 1995, p. 448).

Quando $\varrho(x) = x$, os resultados acima evidenciam a equivalência entre as propriedades assintóticas dos estimadores clássico e robusto, sugerindo que as propriedades do periodograma clássico podem ser estendidas de forma natural para o periodograma tipo M. As propriedades do periodograma clássico têm sido amplamente estudadas em processos estacionários com propriedade de memória curta e longa (ver e.g. Hannan (1973), Yajima (1989), Hurvich & Beltrão (1993) e Robinson (1995)).

3.2. Periodogramas tipo L. Utilizando a expressão apresentada na Eq. 5 sugere-se um estimador robusto para a densidade espectral de processos estacionários, definido por

$$\hat{f}_{rob}(\lambda_j) = \frac{1}{2\pi} \sum_{|h| \leq r} \kappa \left(\frac{h}{r} \right) \tilde{R}(h) e^{ih\lambda_j}, \quad (9)$$

onde $\lambda_j = \frac{2\pi j}{n}$ com $j = 1, 2, \dots, \lfloor \frac{n}{2} \rfloor$, $r = n^\beta$, $0 < \beta < 1$, e n o número de observações. A função $\tilde{R}(\cdot)$ representa um estimador robusto para a autocovariância, sugerido por Ma & Genton (2000), e dado por

$$\tilde{R}(h) = \frac{1}{4} [Q_{n-h}^2(\mathbf{u} + \mathbf{v}) - Q_{n-h}^2(\mathbf{u} - \mathbf{v})], \quad (10)$$

onde \mathbf{u} e \mathbf{v} são vetores que constam das primeiras e últimas $n - h$ observações, respectivamente. O estimador $Q_n(\cdot)$ calcula a k -ésima estatística de ordem das $\binom{n}{2}$ distâncias $\{|z_i - z_j|, i < j\}$, i.e. $Q_n(z) = c_\Phi \cdot \{|z_i - z_j|; i < j\}_{(k)}$, onde $z = (z_1, z_2, \dots, z_n)'$ representa o vetor de dados, c_Φ é uma constante para garantir a propriedade de consistência ($c_\Phi = 2.2191$ para distribuição normal) e $k = \left\lfloor \frac{\binom{n}{2} + 2}{4} \right\rfloor + 1$.

O estimador definido na Eq. 9, quando aplicada a *janela truncada*, é sugerido por Fajardo et al. (2009). Os autores mostram que o *bandwidth* pode ser dado por $r = n^\beta$, onde $0 < \beta < 1$. O valor β que define o *bandwidth* pode ser obtido através de simulações de Monte Carlo utilizando critérios estatísticos de seleção. Fajardo (2011) sugere a aplicação de *janelas* alternativas para melhorar as propriedades estatísticas do estimador. O autor sugere a aplicação das *janelas* de Parzen e Bartlett resulta uma alternativa razoável quando os processos são contaminados por dados atípicos de tipo aditivo. Os resultados empíricos apresentados por Fajardo (2011) sugerem que o periodograma robusto alisado pode ser considerado uma alternativa robusta para estimação do espectro de processos estacionários sob contaminação.

3.2.1. *Medidas de robustez dos estimadores.* A vantagem prática do periodograma robusto se evidencia quando aplicado em séries contaminadas por dados atípicos. Como mostrado por Fajardo et al. (2009), o periodograma clássico é afetado significativamente pela presença de pelo menos um valor atípico na série temporal, i.e. o *ponto de quebra* (PQ) do periodograma clássico é zero. O PQ denotado por γ^* é dado por

$$\gamma^*(T_n, y_1, \dots, y_n) := \max \left\{ \frac{m}{n} : \max_{\mathbb{I}_m} \sup_{v_1, \dots, v_m} |T_n(z_1, \dots, z_n)| < \infty \right\},$$

onde z_1, \dots, z_n é uma amostra obtida substituindo m dados pontuais y_{i_1}, \dots, y_{i_m} por valores arbitrários v_1, \dots, v_m e $\mathbb{I}_m = \{i_1, \dots, i_m\} \subset \{1, \dots, n\}$, i.e. o PQ representa a máxima quantidade de dados atípicos contida no conjunto de dados para que o estimador T_n forneça estimativas razoáveis do parâmetro θ . Maronna (1976) mostra que $\gamma^*(T_n, y_1, \dots, y_n) \leq 0.25$, onde a igualdade é atingida para o estimador L_1 .

O estimador T_n , obtido a partir da Eq. 7, mantém as propriedades de robustez dos M-estimadores para modelos de regressão, amplamente analisados por Hampel et al. (1986), Huber (2004) e Maronna et al. (2006), entre outros. Deste modo, seja $T(F)$ um M-funcional associado ao estimador definido na Eq. 8, e solução de $\int \varrho(e) \mathbf{x} dF = 0$, onde $e := y - \mathbf{x}'T(F)$ e $\sigma = 1$. A função de influência de $T(F)$, para uma certa função de distribuição simétrica F , é dada por

$$\text{FI}(y, T, F) = \varrho(e) \left[\int \frac{\partial}{\partial e} \varrho(e) dF \right]^{-1}.$$

As medidas de robustez para o estimador sugerido na Eq. 9 estão relacionadas às propriedades do estimador $\tilde{R}(\cdot)$. As propriedades estatísticas da função $\tilde{R}(\cdot)$ são provadas por Lévy-Leduc et al. (2011a) e Ma & Genton (2000). Os autores mostram que, para $0 \leq h < n$, a função $\tilde{R}(h)$ satisfaz o teorema central do limite e ponto de quebra temporal menor a 25% (para detalhes ver Ma & Genton (2000)).

4. RESULTADOS EMPÍRICOS

4.1. **Experimentos de Monte Carlo.** O comportamento empírico dos periodogramas robustos é analisado através de simulações de Monte Carlo. As séries observadas foram geradas de um processo $\{z_t; t = 1, 2, \dots, n\}$ com representação dada por

$$z_t = y_t + v_t, \quad (11)$$

onde a sequência $\{v_t; t = 1, 2, \dots, n\}$ é uma sequência de variáveis aleatórias independentes e identicamente distribuídas com distribuição de misturas de normais, dada por $F_v = (1 - \gamma)\delta_0 + \gamma\mathcal{N}(0, \sigma_v^2)$, onde δ_0 representa a distribuição de massa pontual concentrada em zero e $\sigma_v = 10$ e $0 < \gamma < 1$ (ver, e.g. Denby & Martin (1979)). O processo não-contaminado $\{y_t; t = 1, 2, \dots, n\}$ é gerado a partir de um modelo ARFIMA(p, d, q) gaussiano dado por

$$y_t = (1 - B)^{-d} u_t, \quad t = 1, 2, \dots, n, \quad (12)$$

onde o filtro de diferenciação fracionária $(1 - B)^d$, para $d \in \mathbb{R}$, é definido pela expansão binomial $(1 - B)^d = \sum_{j=0}^{\infty} \pi_j B^j$, onde $\pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}$ e $\Gamma(\cdot)$ é a função Gamma. O processo $\{u_t\}$ com representação ARMA(p, q) apresenta ruído branco gaussiano com média 0 e variância 1. O espectro do processo $\{y_t; t = 1, 2, \dots, n\}$ é dado por

$$f_y(\lambda) = \frac{1}{2\pi} \frac{|\Theta(e^{-i\lambda})|^2}{|\Phi(e^{-i\lambda})|^2} \left\{ 2 \operatorname{sen} \left(\frac{\lambda}{2} \right) \right\}^{-2d},$$

onde $\Phi(x) = 1 - \phi_1 x - \dots - \phi_p x^p$ e $\Theta(x) = 1 - \theta_1 x - \dots - \theta_q x^q$. Para a geração de dados com representação dada na Eq. 12, segue-se a metodologia sugerida por Hosking (1981) para valores $d = 0, 0.3$ e $p = q = 1$ com $\phi = 0.5$ e $\theta = 0.3$. Foram obtidas 10000 realizações de tamanho 300 para o processo contaminado com outliers de tipo aditivo, definido na Eq. 11, para valores $\gamma = 0, 0.05, 0.1, 0.25$. As simulações foram realizadas na linguagem de programação matricial **Ox** (ver <http://www.doornik.com>).

Periodogramas tipo M. Os periodogramas calculados para cada realização do processo foram obtidos a partir do estimador $\tilde{f}(\lambda_j) = \frac{1}{2\pi} \tilde{I}(\lambda_j)$, com $0 < j < \lfloor \frac{n}{2} \rfloor$ e $\tilde{I}(\lambda_j)$ definido na Eq. 6. Os periodogramas analisados são apresentados a seguir:

- (1) Periodograma clássico: denotado por *LS* e definido na Eq. 3.
- (2) Periodograma L_1 : denotado por L_1 e calculado através da função $\varrho = \operatorname{sgn}(x)$.
- (3) Periodograma Bisquare: denotado por *BS* e obtido a partir de

$$\varrho(x) = 0.2734x [1 - 0.0456x^2]^2 \text{ se } |x| \leq 4.6850.$$

- (4) Periodograma Hampel: denotado por *HP* e obtido a partir de

$$\varrho(x) = \begin{cases} x, & \text{se } 0 \leq |x| \leq 2; \\ 2 \operatorname{sgn}(x), & \text{se } 2 \leq |x| \leq 4; \\ \frac{8-|x|}{2} \operatorname{sgn}(x), & \text{se } 4 \leq |x| \leq 8; \\ 0, & \text{se } 8 \leq |x| \end{cases}$$

- (5) Periodograma Huber: denotado por *HU* e obtido por

$$\varrho(x) = \begin{cases} x, & \text{if } |x| \leq 1.345; \\ 1.345 \operatorname{sgn}(x), & \text{if } |x| > 1.345. \end{cases}$$

Na Figura 1 são apresentados o espectro teórico e os periodogramas tipo *M* correspondentes a um processo ARMA(1,1) com $\phi = 0.5$ e $\theta = 0.3$ com diferentes níveis de contaminação. Neste cenário se evidencia o baixo desempenho do periodograma clássico sob contaminação nos dados. Entre os estimadores tipo *M*, o periodograma L_1 é o que apresenta menor precisão tanto em séries contaminadas quanto em não-contaminadas. Os resultados mostrados na Figura 1 são quantificados na Tabela 1 através de medidas estatísticas de erro de precisão. Os valores em negrito apresentados na tabela representam os menores valores das medidas calculadas para os diferentes periodogramas. As medidas de erro de precisão são definidas a seguir:

Raíz quadrada do erro quadrático médio (REQM):: $\frac{1}{\lfloor \frac{n}{2} \rfloor} \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} (f(\lambda_j) - \tilde{I}(\lambda_j))^2$;

erro absoluto médio (MAE):: $\frac{1}{\lfloor \frac{n}{2} \rfloor} \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} |f(\lambda_j) - \tilde{I}(\lambda_j)|$;

erro percentual absoluto médio (MAPE):: $\frac{1}{\lfloor \frac{n}{2} \rfloor} \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} \left| \frac{f(\lambda_j) - \tilde{I}(\lambda_j)}{f(\lambda_j)} \right| \times 100$.

A eficiência relativa assintótica (ERA) de $\tilde{f}(\lambda_j)$ é obtida a partir do quociente entre as variâncias assintóticas do periodograma clássico e do periodograma robusto, dada por $ERA(\lambda_j) = \frac{k^4 \sigma^4}{\mathbb{E}^2[\varrho^2(\epsilon_1)]}$, onde σ^4 representa a variância assintótica do periodograma clássico (ver, e.g. Maronna et al. (2006)).

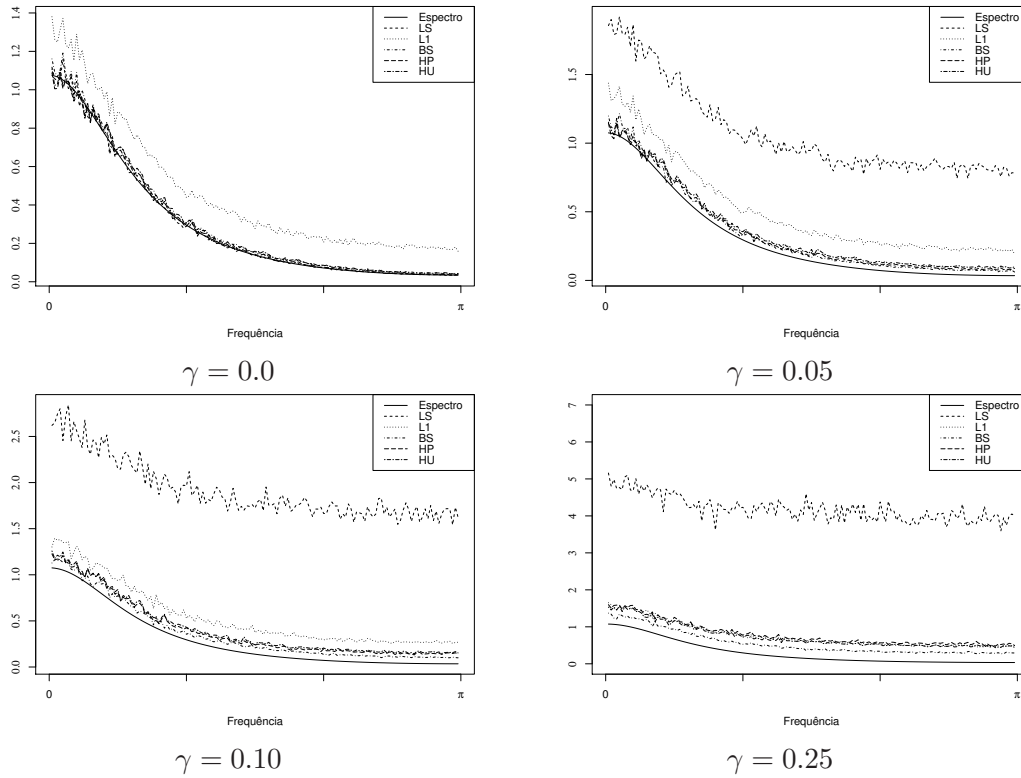


FIGURA 1. Média dos periodogramas tipo M: 10000 realizações de tamanho 300 de um processo ARMA(1,1) com $\phi = 0.5$ e $\theta = 0.3$.

γ	medidas	LS	L_1	BS	HP	HU
0.00	$REQM$	0.0177	0.1766	0.0279	0.0189	0.0247
	MAE	0.0090	0.1728	0.0198	0.0115	0.0193
	$MAPE$	3.2719	162.1418	12.7409	4.9074	12.9720
0.05	$REQM$	0.7808	0.2214	0.0490	0.0526	0.0685
	MAE	0.7795	0.2183	0.0444	0.0503	0.0656
	$MAPE$	830.2500	210.4226	40.1846	51.1762	66.5596
0.10	$REQM$	1.6268	0.2576	0.0762	0.1139	0.1310
	MAE	1.6245	0.2553	0.0737	0.1119	0.1289
	$MAPE$	1741.1820	256.2654	75.6138	117.1856	132.7510
0.25	$REQM$	3.9025	0.4698	0.2596	0.5054	0.4465
	MAE	3.8979	0.4685	0.2586	0.5041	0.4452
	$MAPE$	4177.2350	485.1796	276.0443	534.1787	471.2964

TABELA 1. Medidas de precisão do periodograma tipo M: modelo ARMA(1,1) com $\phi = 0.5$ e $\theta = 0.3$.

Para séries sem contaminação, os periodogramas LS e HP apresentam os menores valores para as medidas de erro de precisão, seguidos pelos periodogramas BS e HU que podem ser considerados alternativas razoáveis para estimação do espectro em séries estacionárias. Em séries contaminadas, os resultados verificam o baixo desempenho do periodograma clássico, sugerindo que o uso de tal estimador deve ser evitado para as análises estatísticas quando existe suspeita de contaminação nas observações. Observe-se que com uma percentagem de contaminação de 5%, o MAPE do estimador LS evidencia um aumento de pouco mais de 250 vezes o valor original, i.e $MAPE = 3.2719$. Em

outras palavras, o resultado sugere que o ponto de quebra amostral para o estimador clássico do espectro é zero.

A sensibilidade mostrada pelo estimador LS em séries contaminadas confirma os resultados obtidos por Fajardo et al. (2009) e Haldrup & Nielsen (2007) para o estimador usual do espectro de processos contaminados. Os autores mostram que a presença de um valor atípico no conjunto de observações é suficiente para alterar as propriedades estatísticas do estimador.

O periodograma L_1 apresenta os maiores valores para os erros de precisão e a menor eficiência relativa entre os periodogramas tipo M aqui estudados (ver Figura 2), mesmo quando as séries não são contaminadas. Para o estimador L_1 os valores das medidas de precisão não sofrem mudanças significativas quando aumenta a porcentagem de contaminação nos dados (ver Tabela 1); por exemplo, para um aumento de 5% de contaminação o MAPE do estimador L_1 evidencia um aumento de 1.3 vezes seu valor inicial (162.1418), enquanto que para os periodogramas BS , HP e HU o aumento ultrapassa pelo menos três vezes o valor do MAPE em séries não contaminadas.

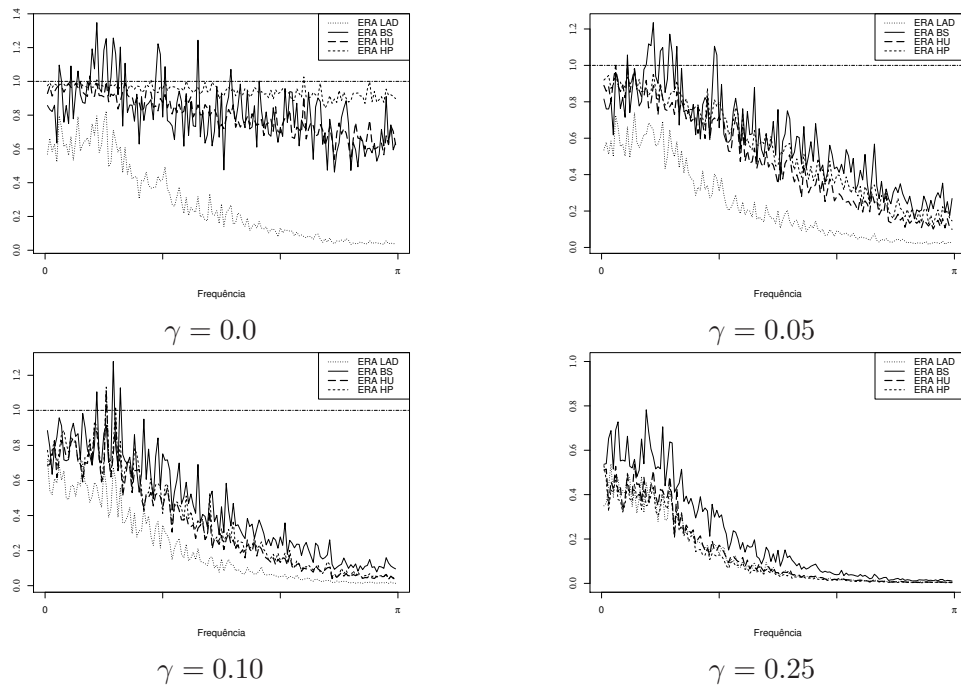


FIGURA 2. ERA dos periodogramas tipo M: modelo ARMA(1,1) com $\phi = 0.5$ e $\theta = 0.3$.

Os resultados empíricos sugerem que o ponto de quebra do estimador L_1 tende a ser superior ao de todos os estimadores aqui analisados. Katkovnik (1998) e Li (2008) sugerem a aplicação do periodograma L_1 no contexto de eventos extremos. Os autores mostram através de simulações de Monte Carlo que o periodograma L_1 pode ser considerado uma alternativa razoável para estimação do espectro de processos com distribuições de caudas pesadas.

Sob as medidas calculadas na Tabela 1, o periodograma BS apresenta melhor desempenho na estimação do espectro de processos contaminados, ainda o estimador HP mostra-se como o segundo estimador com melhor desempenho na presença de dados atípicos nos dados. O resultados indicam que a ERA dos estimadores BS , HU e HP é maior que 80% em ausência de contaminação nos dados, apresentando maior eficiência o estimador HP com uma porcentagem próxima de 95%. Observe-se que para frequências próximas de π , os estimadores apresentam um aumento na variância, tornando-os menos eficientes mesmo para dados não-contaminados. Em séries contaminadas com $\gamma = 0.25$, se evidencia uma diminuição na eficiência dos estimadores, sugerindo a necessidade do desenvolvimento de alternativas mais eficientes nesses cenários.

Os resultados na Tabela 1 confirmam os resultados apresentados por Hampel et al. (1986) sobre o ponto de quebra amostral para os estimadores tipo M .

O estudo empírico é estendido para realizações do processo ARFIMA(0,0.3,0). Neste contexto, os periodogramas tipo M apresentam maior precisão para estimação da densidade espectral do processo, quando considerados diferentes níveis de contaminação nos dados. Os resultados mostrados nas Tabelas 1 e 2 sugerem que os estimadores tipo M apresentam um melhor desempenho na estimação do espectro de processos com propriedade de memória longa que para o caso de processos com propriedade de memória curta.

γ	medidas	LS	L_1	BS	HP	HU
0.00	$REQM$	0.0185	0.1228	0.0227	0.0114	0.0164
	MAE	0.0085	0.1211	0.0150	0.0073	0.0134
	$MAPE$	3.9154	83.0091	8.6837	4.1895	8.4438
0.05	$REQM$	0.7759	0.1530	0.0356	0.0356	0.0500
	MAE	0.7742	0.1514	0.0310	0.0341	0.0487
	$MAPE$	538.6705	103.9826	20.2744	23.3259	33.6986
0.10	$REQM$	1.5635	0.1881	0.0527	0.0799	0.0985
	MAE	1.5616	0.1874	0.0514	0.0778	0.0969
	$MAPE$	1089.9330	129.5482	35.2686	53.4477	66.7723
0.25	$REQM$	4.0600	0.3372	0.1757	0.3544	0.3388
	MAE	4.0561	0.3366	0.1750	0.3532	0.3375
	$MAPE$	2838.0410	233.9922	121.9709	246.1297	235.1634

TABELA 2. Medidas de precisão Periodograma tipo M: modelo ARFIMA(0,0.3,0)

A Figura 3 mostra o comportamento empírico dos estimadores com respeito ao espectro teórico do processo ARFIMA(0,0.3,0). Observa-se que o periodograma BS destaca-se por apresentar o melhor desempenho quando as séries são contaminadas por dados atípicos. O segundo melhor desempenho para todos os níveis de contaminação analisados é apresentado pelo periodograma HP . Mesmo para séries não contaminadas, o estimador apresenta-se como uma alternativa razoável para estimação do espectro.

A eficiência relativa dos periodogramas BS , HP e HU é maior que 85% quando as séries não apresentam dados atípicos, sendo o periodograma HP o mais eficiente com uma porcentagem de aproximadamente 95%. No cenário contaminado, a ERA dos estimadores evidencia uma diminuição em função da porcentagem de contaminação das observações (ver Figura 4). Observe-se que os estimadores sugeridos podem ser considerados alternativas razoáveis para estimação da densidade espectral de processos estacionários, mesmo quando as séries não apresentam contaminação.

Entre os estimadores analisados, o periodograma L_1 mostra o pior desempenho quanto a precisão e eficiência relativa, porém resulta o estimador com maior ponto de quebra dentre os estimadores tipo M aqui analisados. Na prática, o uso do estimador L_1 deve ser evitado para as análises de séries estacionárias na presença de outliers de tipo aditivo.

Periodogramas tipo L. A seguir são apresentados os resultados dos experimentos de Monte Carlo para analisar o comportamento do estimador definido na Eq. 9 em séries contaminadas. Foram calculadas medidas de precisão dos periodogramas alisados com respeito ao espectro teórico dos processos sem contaminação. A eficiência relativa assintótica (ERA) do periodograma robusto é obtida a partir do quociente entre as variâncias do estimador definido na Eq. 1 e o periodograma alisado, obtido a partir da Eq. 9 usando as janelas de Parzen, Tuckey-Hamming (TH), Bartlett e truncada. As janelas são aplicadas com diferentes valores do $r = n^\beta$, onde $\beta = 0.7, 0.8$ e 0.9 .

Os resultados apresentados na Tabela 3 mostram as medidas de precisão correspondentes ao estimador alisado com as diferentes janelas de alisamento, quando o processo sob análise segue uma

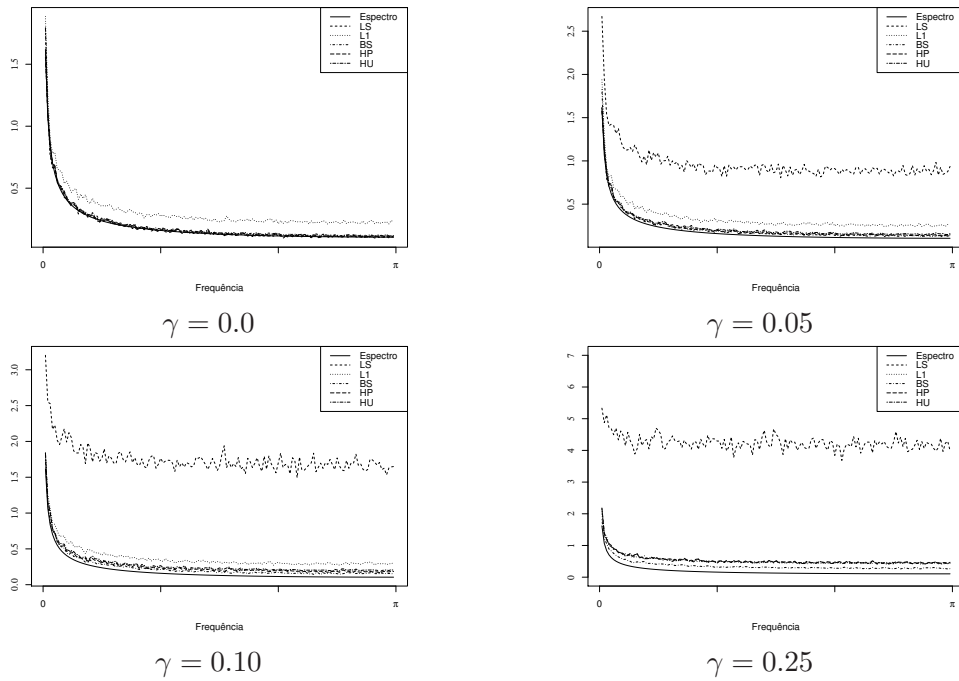


FIGURA 3. Média dos periodogramas tipo M: 10000 realizações de tamanho 300 de um processo ARFIMA(0, 0.3, 0).

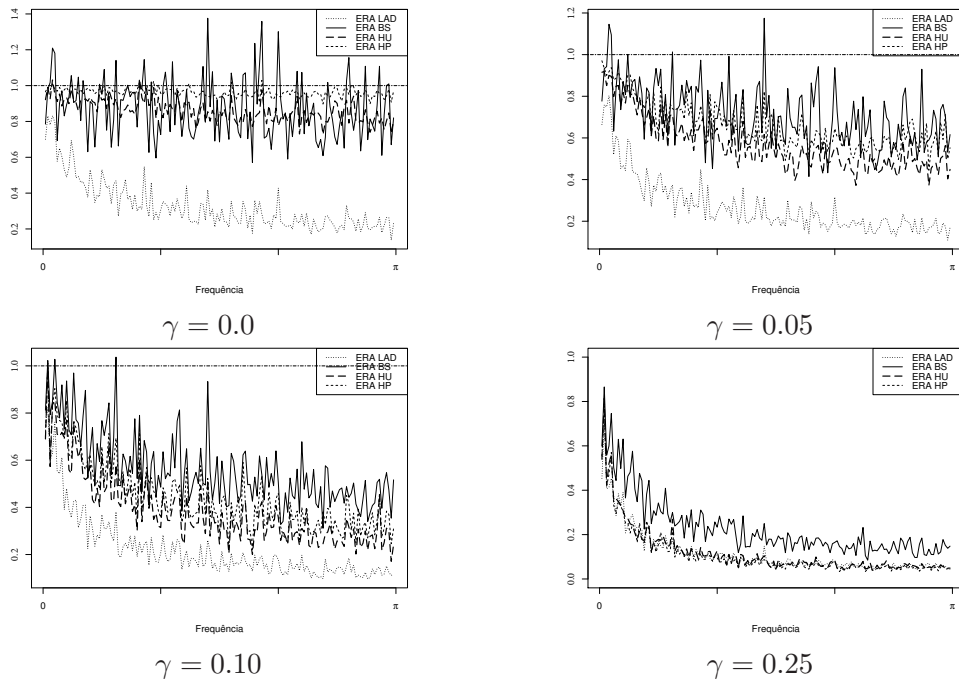


FIGURA 4. Eficiência relativa assintótica dos periodogramas tipo M: modelo ARFIMA(0, 0.3, 0).

representação ARMA(1,1) com $\phi = 0.5$ e $\theta = 0.3$. A aplicação das janelas de Parzen, *TH* e Bartlett apresentam-se como alternativas para melhorar o desempenho do estimador robusto para estimação do espectro de processos ARMA na presença de observações atípicas. Destaca-se o desempenho do estimador quando aplicada a janela de Bartlett, o qual apresenta o melhor desempenho para séries

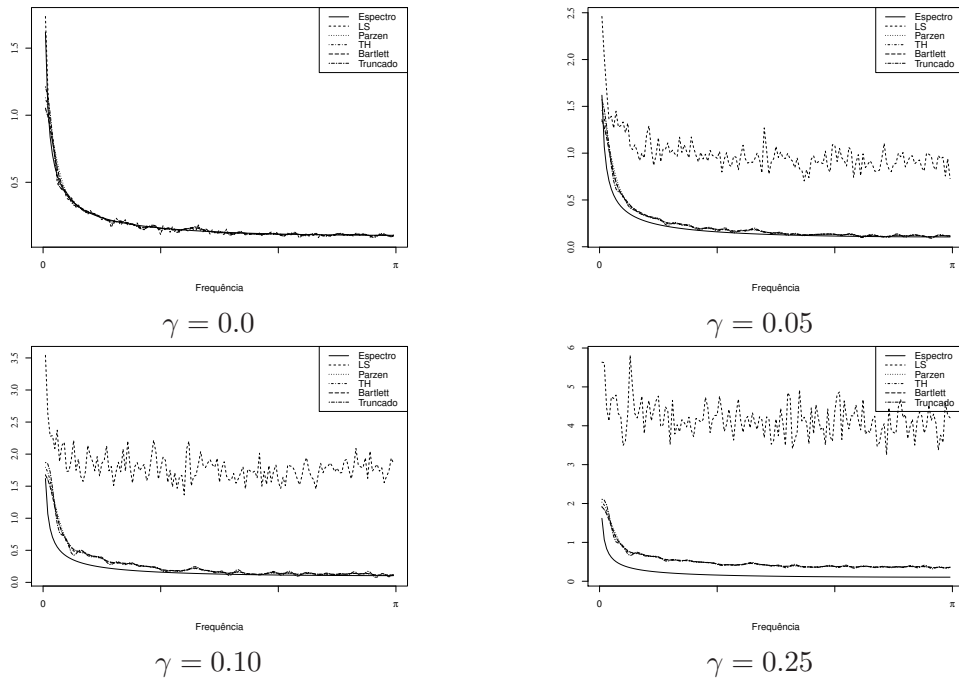


FIGURA 5. Média dos periodogramas tipo L com $\beta = 0.7$: 10000 realizações de tamanho 300 de um processo ARFIMA(0, 0.3, 0).

contaminadas com diferentes níveis de contaminação. Entre os estimadores tipo L , o estimador alisado com a janela de Parzen apresenta o melhor desempenho para séries sem contaminação. A aplicação da janela TH apresenta-se como uma alternativa razoável para estimação do espectro, mesmo quando existe a suspeita de contaminação no conjunto de observações. Os maiores valores nas medidas de precisão são apresentados quando aplicada a janela truncada, a qual evidencia melhor desempenho quando $\beta = 0.7$.

No caso ARFIMA(0, 0.3, 0), o estimador alisado com as janelas TH e truncada apresentam maior precisão quando o valor do parâmetro β é 0.7. O desempenho do estimador quando aplicadas as janelas TH e truncada diminui para os valores $\beta = 0.8$ e 0.9 (ver Tabela 4). As diferenças entre o desempenho do estimador, quando aplicada uma janela de alisamento, não são significativas para valores do *bandwidth* calculados com $\beta = 0.7$ ou $\beta = 0.8$, i.e. não se evidenciam diferenças significativas nas medidas de precisão, quando calculadas para os diferentes valores do parâmetro β , porém o uso da janela truncada não resulta a melhor opção entre as janelas testadas para aplicação no periodograma alisado (ver Tabela 4). Os resultados indicam que as propriedades estatísticas do periodograma robusto alisado melhoram significativamente devido ao uso da janela utilizada.

A resistência dos estimadores, quanto à presença de dados atípicos na série temporal, não depende do tipo de janela aplicada no periodograma e sim das frequências utilizadas no cálculo do mesmo, portanto o ponto de quebra amostral não recebe influência do tipo de janela aplicada no estimador robusto. Empiricamente observa-se que o valor do PQ resulta menor que 10%, pois a precisão dos estimadores resulta seriamente comprometida quando o nível de contaminação aumenta (ver Figuras 5, 6 e 7).

Os valores da ERA são próximos de 65% quando o processo não apresenta contaminação. O estimador com a janela truncada apresenta os menores valores para ERA, sugerindo que a aplicação das janelas de Parzen, TH ou Bartlett pode melhorar a eficiência do estimador robusto. No caso contaminado, as Figuras 8 e 9 evidenciam uma diminuição significativa da eficiência do estimador quando aumenta a porcentagem de contaminação no conjunto de observações.

$\beta = 0.7$					
γ	medidas	<i>Parzen</i>	<i>TH</i>	<i>Bartlett</i>	<i>Truncado</i>
0.00	<i>REQM</i>	0.0212	0.0281	0.0292	0.0440
	<i>MAE</i>	0.0117	0.0150	0.0163	0.0228
	<i>MAPE</i>	6.5394	7.2596	9.9014	9.8748
0.05	<i>REQM</i>	0.0936	0.0995	0.0911	0.1125
	<i>MAE</i>	0.0607	0.0623	0.0584	0.0679
	<i>MAPE</i>	23.9648	24.4735	21.2728	27.1426
0.10	<i>REQM</i>	0.2234	0.2267	0.2173	0.2368
	<i>MAE</i>	0.1476	0.1487	0.1437	0.1519
	<i>MAPE</i>	56.8197	57.3552	53.1758	59.4265
0.25	<i>REQM</i>	0.7888	0.7962	0.7771	0.8200
	<i>MAE</i>	0.5146	0.5167	0.5070	0.5268
	<i>MAPE</i>	172.5879	173.2934	165.1584	181.3773
$\beta = 0.8$					
p	medidas	<i>Parzen</i>	<i>TH</i>	<i>Bartlett</i>	<i>Truncado</i>
0.00	<i>REQM</i>	0.0304	0.0354	0.0328	0.0487
	<i>MAE</i>	0.0149	0.0174	0.0156	0.0258
	<i>MAPE</i>	6.0567	7.3314	6.9163	11.1520
0.05	<i>REQM</i>	0.1014	0.1063	0.0993	0.1200
	<i>MAE</i>	0.0631	0.0652	0.0623	0.0702
	<i>MAPE</i>	24.6493	25.3046	23.3842	28.2097
0.10	<i>REQM</i>	0.2280	0.2325	0.2256	0.2478
	<i>MAE</i>	0.1491	0.1505	0.1472	0.1557
	<i>MAPE</i>	57.5332	58.1971	55.5902	61.2707
0.25	<i>REQM</i>	0.7993	0.8103	0.7963	0.8439
	<i>MAE</i>	0.5176	0.5214	0.5146	0.5350
	<i>MAPE</i>	173.3536	175.2932	170.0304	185.2228
$\beta = 0.9$					
p	medidas	<i>Parzen</i>	<i>TH</i>	<i>Bartlett</i>	<i>Truncado</i>
0.00	<i>REQM</i>	0.0357	0.0393	0.0353	0.0644
	<i>MAE</i>	0.0181	0.0216	0.0195	0.0370
	<i>MAPE</i>	8.0385	9.6417	9.2577	15.2175
0.05	<i>REQM</i>	0.4539	0.4558	0.4542	0.4630
	<i>MAE</i>	0.4439	0.4448	0.4446	0.4461
	<i>MAPE</i>	416.6903	416.7427	418.0991	416.9504
0.10	<i>REQM</i>	0.4867	0.4898	0.4869	0.4981
	<i>MAE</i>	0.4436	0.4447	0.4443	0.4469
	<i>MAPE</i>	335.5855	335.5932	337.3849	335.9566
0.25	<i>REQM</i>	0.8150	0.8283	0.8156	0.8658
	<i>MAE</i>	0.5231	0.5282	0.5227	0.5444
	<i>MAPE</i>	176.2388	179.0922	174.7544	197.0127

TABELA 3. Medidas de precisão do periodograma tipo L : modelo ARMA(1,1) com $\phi = 0.5$ e $\theta = 0.3$.

5. COMENTÁRIOS FINAIS

Neste artigo estudam-se estimadores robustos para o espectro de processos estacionários contaminados com outliers de tipo aditivo. Os estimadores são classificados em dois grupos. O

$\beta = 0.7$					
γ	medidas	<i>Parzen</i>	<i>TH</i>	<i>Bartlett</i>	<i>Truncado</i>
0.00	<i>REQM</i>	0.0522	0.0446	0.0488	0.0403
	<i>MAE</i>	0.0120	0.0111	0.0109	0.0139
	<i>MAPE</i>	3.4747	3.9750	3.5849	6.1065
0.05	<i>REQM</i>	0.0767	0.0734	0.0693	0.0730
	<i>MAE</i>	0.0405	0.0385	0.0401	0.0377
	<i>MAPE</i>	16.1522	15.7194	17.0576	16.3730
0.10	<i>REQM</i>	0.1430	0.1453	0.1361	0.1524
	<i>MAE</i>	0.0790	0.0785	0.0789	0.0801
	<i>MAPE</i>	32.8495	32.3349	34.0494	33.5176
0.25	<i>REQM</i>	0.3405	0.3412	0.3376	0.3441
	<i>MAE</i>	0.3189	0.3185	0.3189	0.3183
	<i>MAPE</i>	200.4568	199.9444	201.6625	199.5148
$\beta = 0.8$					
γ	medidas	<i>Parzen</i>	<i>TH</i>	<i>Bartlett</i>	<i>Truncado</i>
0.00	<i>REQM</i>	0.0433	0.0412	0.0440	0.0505
	<i>MAE</i>	0.0112	0.0124	0.0113	0.0218
	<i>MAPE</i>	4.2429	5.0965	4.4831	9.3537
0.05	<i>REQM</i>	0.0730	0.0742	0.0709	0.0912
	<i>MAE</i>	0.0383	0.0381	0.0391	0.0412
	<i>MAPE</i>	15.7755	15.9503	16.5413	18.0506
0.10	<i>REQM</i>	0.1460	0.1498	0.1429	0.1630
	<i>MAE</i>	0.0785	0.0789	0.0790	0.0839
	<i>MAPE</i>	32.2739	32.3974	33.2495	35.8455
0.25	<i>REQM</i>	0.5165	0.5268	0.5123	0.5823
	<i>MAE</i>	0.3134	0.3147	0.3148	0.3235
	<i>MAPE</i>	136.4854	136.8104	138.5574	140.4886
$\beta = 0.9$					
γ	medidas	<i>Parzen</i>	<i>TH</i>	<i>Bartlett</i>	<i>Truncado</i>
0.00	<i>REQM</i>	0.0402	0.0376	0.0376	0.0354
	<i>MAE</i>	0.0129	0.0148	0.0135	0.0201
	<i>MAPE</i>	5.4695	6.7888	6.0912	11.1038
0.05	<i>REQM</i>	0.0750	0.0788	0.0745	0.0894
	<i>MAE</i>	0.0380	0.0390	0.0384	0.0456
	<i>MAPE</i>	16.0405	16.6804	16.6406	20.6241
0.10	<i>REQM</i>	0.1513	0.1558	0.1500	0.1676
	<i>MAE</i>	0.0795	0.0816	0.0810	0.0892
	<i>MAPE</i>	32.7155	33.8761	33.8776	39.2540
0.25	<i>REQM</i>	0.5340	0.5581	0.5441	0.6149
	<i>MAE</i>	0.3160	0.3198	0.3189	0.3360
	<i>MAPE</i>	137.0726	137.8391	138.3392	146.7352

TABELA 4. Medidas de precisão do periodograma tipo L: processo ARFIMA(0, 0.3, 0).

primeiro grupo baseado em M-estimadores dos parâmetros da representação espectral de processos estacionários, aplicando as funções *bisquare*, Hampel e Huber. O segundo grupo baseado em L-estimadores da função de autocovariância de processos estacionários na Eq. 9 para as janelas de Parzen, Tuckey-Hamming e Bartlett. Os comportamentos dos estimadores foram comparados através de simulações de Monte Carlo, indicando a superioridade do estimador robusto baseado em M-estimadores quanto a precisão e ERA. Os periodogramas *bisquare* e Hampel se destacam por apresentar maior eficiência relativa e maior precisão com respeito à densidade espectral do processo. O periodograma L_1 apresenta menor aumento nas medidas de erro de precisão quando o nível de contaminação aumenta, sugerindo maior resistência à presença de dados atípicos para

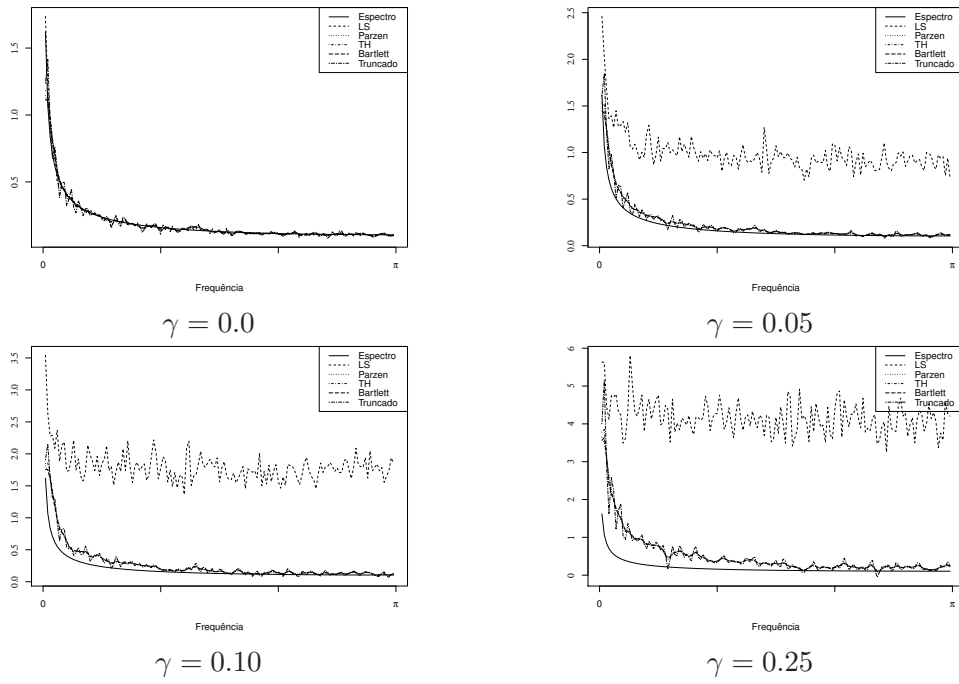


FIGURA 6. Média dos periodogramas tipo L com $\beta = 0.8$: 10000 realizações de tamanho 300 de um processo ARFIMA(0, 0.3, 0).

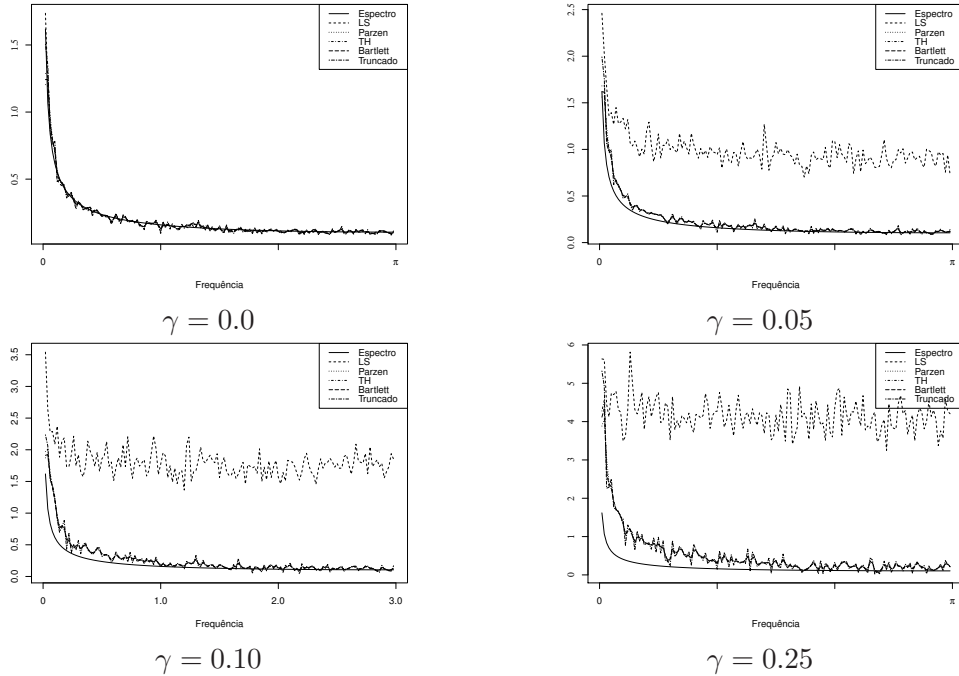


FIGURA 7. Média dos periodogramas tipo L com $\beta = 0.9$: 10000 realizações de tamanho 300 de um processo ARFIMA(0, 0.3, 0).

diferentes níveis de contaminação, porém o periodograma L_1 apresenta o pior desempenho entre os estimadores analisados.

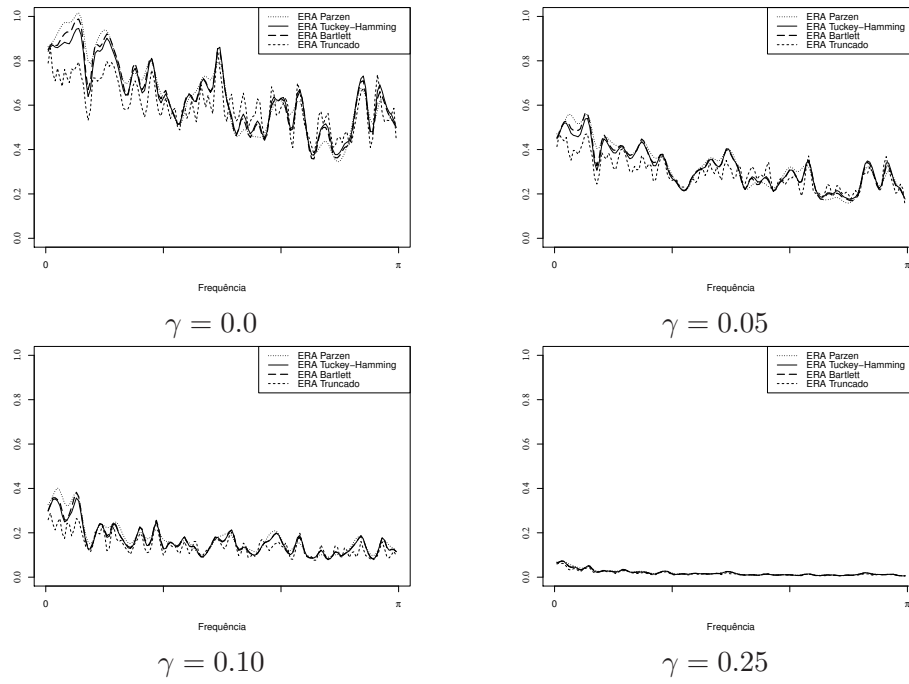


FIGURA 8. Eficiência relativa assintótica do periodograma tipo L com $\beta = 0.7$: processo ARFIMA(0, 0.3, 0).

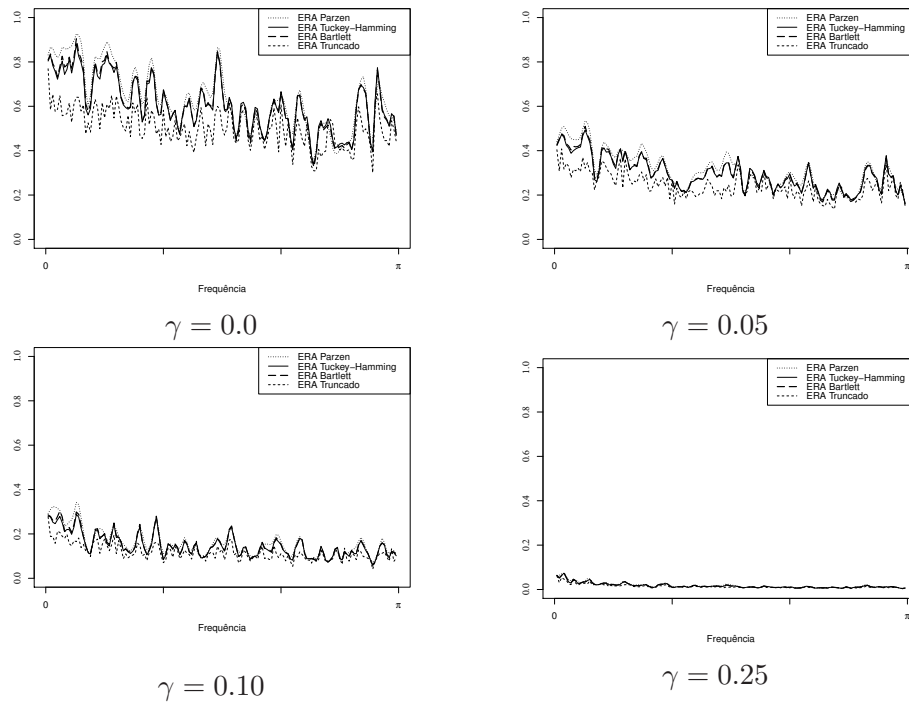


FIGURA 9. Eficiência relativa assintótica do periodograma tipo L com $\beta = 0.8$: processo ARFIMA(0, 0.3, 0).

Através de experimentos de Monte Carlo mostra-se que a utilização de janelas alternativas permite melhorar as propriedades estatísticas do periodograma robusto truncado sugerido por Fajardo et al. (2009). Para aplicação das janelas sugerem-se os valores $\beta = 0.7$ para as janelas de TH e Bartlett e o valor $\beta = 0.9$ para a janela de Parzen dado que, de acordo com os resultados empíricos, esses valores permitem melhorar a eficiência assintótica e a precisão dos estimadores no cálculo de estimativas do espectro de processos estacionários. Porém, as propriedades estatísticas dos estimadores, assim como a definição de um *bandwidth* ótimo no caso contaminado, continua sendo um problema aberto. Recentemente, os trabalhos desenvolvidos por Lévy-Leduc et al. (2011a) e Lévy-Leduc et al. (2011b) mostram que o estimador robusto da função de autocovariância, sugerido por Ma & Genton (2000), satisfaz o teorema central do limite no contexto de processos estacionários, contudo não existem resultados teóricos que permitam estabelecer as propriedades assintóticas do estimador robusto do espectro.

Finalmente, a classe de M-estimadores destaca-se por apresentar estimadores com maior eficiência relativa e precisão, assim como menor custo computacional para o cálculo de estimativas, se comparado com o tempo de cálculo dos L-estimadores. As propriedades de robustez dos estimadores tipo M são uma extensão natural da teoria de M-estimadores para análise de regressão linear. Cabe ressaltar ainda que o ponto de quebra dos periodogramas tipo M não é influenciado pela posição do dado atípico ao longo da série. Estimadores baseados em um estimador robusto da autocovariância apresentam maior dificuldade para provar as propriedades de robustez dos mesmos, devido a que o efeito de um dado atípico na função de autocovariância é diferente dependendo da posição do mesmo ao longo da série.

REFERÊNCIAS

- Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, John Wiley & Sons.
- Arcones, M. (1998), ‘Asymptotic theory for M-estimates over a convex kernel’, *Econometric Theory* **14**, 387–422.
- Arcones, M. (2001), ‘Asymptotic distribution of regression M-estimators’, *Journal of statistical planning and inference* **97**, 235–261.
- Bai, Z. D., Rao, C. R. & Wu, Y. (1992), ‘M-estimates of multivariate linear regression parameters under a convex discrepancy function’, *Statistica sinica* **2**, 237–254.
- Bai, Z. D., Rao, C. R. & Wu, Y. H. (1997), M-estimation of multivariate linear regression by minimizing the difference of two convex functions, in ‘Handbook of Statistics’, Vol. 15, Elsevier Science, North-Holland, pp. 1–19.
- Beltrão, K. & Bloomfield, P. (1987), ‘Determining the bandwidth of a kernel spectrum estimate’, *Journal of Time Series Analysis* **8**, 21–38.
- Brillinger, D. (1981), *Time Series: Data Analysis and Theory*, Siam.
- Brockwell, P. J. & Davis, R. A. (2006), *Time Series: Theory and Methods*, second edn, Springer Series in Statistics.
- Chen, Z. & Hannan, E. J. (1980), ‘The distribution of periodogram ordinates’, *Journal of time series analysis* **1**, 73–82.
- Denby, L. & Martin, R. D. (1979), ‘robust estimation of the first order autoregressive parameter’, *Journal of the American Statistical Association* **74**, 140–146.
- Fajardo, F. (2011), ‘On some alternatives for robust estimation of the spectrum in stationary processes’, *Brazilian review of econometrics* p. Forthcoming.
- Fajardo, F., Reisen, V. A. & Cribari-Neto, F. (2009), ‘Robust estimation in long-memory processes under additive outliers’, *Journal of Statistical Planning and Inference* **139**, 2511–2525.
- Fox, A. J. (1972), ‘Outliers in time series’, *Journal of the Royal Statistical Society* **34**(B), 350–363.
- Fuller, W. A. (1996), *Introduction to statistical time series*, John Wiley & Sons.
- Geweke, J. & Porter-Hudak, S. (1983), ‘The estimation and application of long memory time series model’, *Journal of Time Series Analysis* **4**, 221–238.

- Haberman, S. J. (1989), ‘Concavity and estimation’, *The Annals of Statistics* **17**, 1631–1661.
- Haldrup, N. & Nielsen, M. O. (2007), ‘Estimation of fractional integration in the presence of data noise’, *Computational Statistics & Data Analysis* **51**, 3100–3114.
- Hampel, F., Ronchetti, E., Rousseeuw, P. & Stahel, W. (1986), *Robust statistics: The approach based on influence functions*, John Wiley & sons.
- Hannan, E. J. (1973), ‘Central limit theorems for time series regression’, *Probability theory and related fields* **26**, 157–170.
- Hosking, J. R. (1981), ‘Fractional differencing’, *Biometrika* **68**, 165–176.
- Huber, P. (1964), ‘Robust estimation of a location parameter’, *The Annals of Mathematical Statistics* **35**, 73–101.
- Huber, P. J. (2004), *Robust Statistics*, third edn, John Wiley & Sons.
- Hurvich, C. & Beltrão, K. (1990), ‘Cross-validatory choice of a spectrum estimate and its connections with aic’, *Journal of Time Series Analysis* **11**, 121–137.
- Hurvich, C. M. & Beltrão, K. I. (1993), ‘Asymptotics for low-frequency ordinates of the periodogram of a long-memory time series’, *Journal of Time Series Analysis* **14**(5), 455–472.
- Johnson, N., Kotz, S. & Balakrishnan, N. (1995), *Continuos univariate distributions*, Vol. 2, second edn, John Wiley & Sons.
- Katkovnik, V. (1998), ‘Robust M-periodogram’, *IEEE Transactions on signal processing* **46**, 3104–3109.
- Lévy-Leduc, C., Boistard, H., Moulines, E., Taqqu, M. & Reisen, V. A. (2011a), ‘Asymptotic properties of U-processes under long-range dependence’, *The annals of statistics* **39**, 1399–1426.
- Lévy-Leduc, C., Boistard, H., Moulines, E., Taqqu, M. & Reisen, V. A. (2011b), ‘Robust estimation of the scale and the autocovariance function of gaussian short and long-range dependent processes’, *Journal of Time Series Analysis* **32**, 135–156.
- Li, T. (2008), ‘Laplace periodogram for time series analysis’, *Journal of the American Statistical Association* **103**, 757–768.
- Ma, Y. & Genton, M. (2000), ‘Highly robust estimation of the autocovariance function’, *Journal of Time Series Analysis* **21**, 663–684.
- Maronna, R. (1976), ‘Robust M-estimators of multivariate location and scatter’, *The annals of statistics* **4**, 51–67.
- Maronna, R., Martin, R. D. & Yohai, V. (2006), *Robust statistics*, John Wiley & sons.
- Martin, R. D. & Thomson, D. (1982), ‘Robust-resistant spectrum estimation’, *Proceedings of the IEEE* **70**(9), 1097–1115.
- Priestley, M. B. (1981), *Spectral Analysis and Time Series*, Academic Press.
- Robinson, P. (1996), ‘Optimal spectral bandwidth for long memory’, *Statistica sinica* **6**, 97–112.
- Robinson, P. M. (1995), ‘Log-periodogram regression of time series with long range dependence’, *The annals of statistics* **23**, 1048–1072.
- Rockafellar, R. T. (1981), *Theory of subgradients and Its applications to problems of optimization*, Research and education in mathematics, Heldermann Verlag Berlin.
- Rousseeuw, P. J. & Croux, C. (1993), ‘Alternatives to the median absolute deviation’, *Journal of the American Statistical Association* **88**, 1273–1283.
- Serfling, R. (1980), *Approximation theorems of mathematical statistics*, John Wiley & sons.
- Spangl, B. & Dutter, R. (2005), ‘On robust estimation of power spectra’, *Austrian Journal of Statistics* **34**(2), 199–210.
- Taniguchi, M. (1979), ‘On estimation of parameters of gaussian stationary processes’, *Journal of applied Probability* **16**, 575–591.
- Tatum, L. & Hurvich, C. (1993), ‘High breakdown methods of time series analysis’, *Journal of the Royal Statistical Society B* **55**, 881–896.
- Yajima, Y. (1989), ‘A central limit theorem for fourier transform of strongly dependent stationary’, *journal of time series analysis* **10**, 375–383.

Yohai, V. & Maronna, R. (1979), 'Asymptotic behavior of M-estimators for the linear model', *The annals of statistics* **7**, 258–268.

ESTIMADORES ROBUSTOS PARA O PARÂMETRO DE DIFERENCIAÇÃO FRAÇIONÁRIA EM PROCESSOS ARFIMA(p, d, q)

FABIO A. FAJARDO[†] - VALDÉRIO A. REISEN

Departamento de Estatística, UFMG, Belo Horizonte/MG, Brasil
Departamento de Estatística, UFES, Vitória/ES, Brasil

RESUMO. Neste trabalho propõe-se um estimador robusto para o parâmetro de diferenciação fracionária de processos ARFIMA. O estimador apresenta-se como uma variante do estimador log-periodograma (LP) sugerido por Geweke & Porter-Hudak (1983), onde o periodograma clássico é substituído por um estimador robusto do espectro, baseado na soma de quadrados dos M-estimadores dos parâmetros da representação espectral do processo. O desempenho do estimador é comparado com o estimador proposto por Fajardo et al. (2009), através de simulações de Monte Carlo.

Palavras-chave. robustez, dados atípicos, periodograma, modelos ARFIMA.

1. INTRODUÇÃO

Na análise de séries econômicas e financeiras é comum encontrar a presença de observações influenciadas por eventos externos que podem provocar mudanças na dinâmica das séries, algumas vezes de forma transitória e outras de forma permanente. Essas observações são conhecidas na literatura como dados atípicos ou *outliers* e, dependendo da sua natureza, seus efeitos podem ser significativos sobre os resultados obtidos.

Estudos baseados na suposição que as séries observadas são geradas por um processo Autorregressivo e de Médias Móveis (ARMA) mostram a influência dos *outliers* nas estimativas dos parâmetros e nas previsões obtidas a partir dos modelos ajustados (ver, e.g., Chang et al. (1988), Chen & Liu (1993), Deutsch et al. (1990) e Chan (1992, 1995) entre outros).

Na década do 80, Granger & Joyeux (1980) e Hosking (1981) apresentaram uma extensão dos processos ARMA integrados (ARIMA) em que o parâmetro de integração assume valores fracionários: o processo ARFIMA. Hosking (1981) provou que séries com representação ARFIMA(p, d, q), com as raízes dos polinômios autorregressivos e de médias móveis fora do círculo unitário, são estacionárias mas não invertíveis quando $d \leq -0.5$. Para $|d| < 0.5$ as séries são estacionárias e invertíveis; e quando $d \geq 0.5$ o processo ARFIMA é não-estacionário. Hosking (1981) mostrou que para valores $d > 0$, as séries apresentam a propriedade de memória longa, sendo esta última caracterizada por correlações estatisticamente significativas entre observações distantes; equivalentemente, a função de densidade espectral tem singularidade na frequência zero.

Na prática, um estimador bastante conhecido para estimação do parâmetro de memória é proposto por Geweke & Porter-Hudak (1983) (LP), obtido através de um modelo de regressão linear do logaritmo da função periodograma. Outras alternativas são propostas por Fox & Taqqu (1986), Künsch (1987), Robinson (1995a, 1995b) entre outros. Hurvich et al. (1998) e Robinson (1995b) provaram a consistência e normalidade assintótica do estimador em processos estacionários. No caso não-estacionário, o estimador LP apresenta propriedades estatísticas diferentes para diferentes

Date: 20 de dezembro de 2011.

Email addresses: ffajardo@est.dout.ufmg.br.

valores do d . Por exemplo, para $d \in [\frac{1}{2}, 1]$ o estimador é consistente, sendo normalmente distribuídos para $d \in [\frac{1}{2}, \frac{3}{4}]$. Para a região $d \in (\frac{3}{4}, 1]$ o estimador não apresenta distribuição normal (cf. Phillips(1999, 2007), Kim & Phillips (2006) e Velasco (1999a, 1999b)).

O estudo de modelos de memória longa na presença de *outliers* tem sido, recentemente, um assunto de muito interesse para pesquisadores da área. Evidências empíricas mostram que as estimativas do parâmetro de memória são significativamente alteradas pela presença de *outliers* do tipo aditivo na série temporal. Haldrup & Nielsen (2007) mostraram, através de simulações de Monte Carlo, que a presença de erros de medição, *outliers* e mudanças estruturais em séries temporais com tamanhos de amostras relativamente pequenos afetam seriamente as estimativas obtidas para o parâmetro de memória causando um viés nas estimativas do parâmetro de integração fracionária. Os autores concluíram que os estimadores semiparamétricos obtidos por regressão apresentam viés relativamente menor quando o *bandwidth*, que corresponde ao número de frequências utilizadas para o cálculo das estimativas, é reduzido. Fajardo et al. (2009) mostraram que o viés pode ser explicado por uma translação da função de densidade espectral do processo observado.

No caso estacionário, várias propostas para estimação robusta do parâmetro d podem ser encontradas na literatura. Por exemplo, Agostinelli & Bisaglia (2003) sugeriram um método paramétrico baseado em verossimilhança ponderada como uma modificação do estimador proposto por Beran (1994). Fajardo et al. (2009) propuseram uma metodologia robusta para estimar os parâmetros do modelo ARFIMA na presença de *outliers* do tipo aditivo. O método proposto pelos autores é baseado em um estimador *plugin* do estimador LP, onde a função periodograma é substituída por um estimador robusto do espectro do processo. Sob a condição de não-estacionariedade na presença de observações atípicas, as propriedades estatísticas dos estimadores são comprometidas; e a estimação através deles não é recomendável para aplicações práticas.

Seguindo a metodologia sugerida por Fajardo et al. (2009), neste trabalho propõe-se um estimador robusto do parâmetro de memória de processos ARFIMA estacionários e não-estacionários. O estimador baseia-se na utilização de um estimador robusto do espectro na equação de regressão que permite o cálculo do estimador LP. O estimador robusto é função dos M-estimadores dos parâmetros na representação espectral do processo.

Resultados empíricos através de simulações de Monte Carlo sugerem que as propriedades estatísticas dos estimadores LP e robusto se mostram equivalentes para tamanhos de amostra suficientemente grandes em séries não-contaminadas. O novo estimador do parâmetro de memória é uma alternativa eficaz quando a série temporal é contaminada por dados atípicos de tipo aditivo. Destaca-se o desempenho do estimador robusto do parâmetro d mesmo quando as séries não são contaminadas, sugerindo que pode ser considerado uma alternativa para estimação do parâmetro de diferenciação fracionária em séries sem contaminação.

Este documento encontra-se organizado da seguinte forma: na Seção 2 apresentam-se alguns conceitos relacionados a processos com parâmetro de memória e estimadores semiparamétricos para o parâmetro de diferenciação fracionária são apresentados. Estimadores robustos para estimação do parâmetro de memória em processos estacionários e não-estacionários são apresentados na Seção 3. A Seção 4 contém os resultados empíricos do estimador proposto e uma aplicação com dados reais é desenvolvida. Finalmente, na Seção 5 são apresentados os principais comentários do trabalho e algumas propostas para o desenvolvimento da pesquisa futura.

2. ESTIMADORES DO PARÂMETRO DE MEMÓRIA

Seja $\{y_t; t = 1, 2, \dots, n\}$ um processo com parâmetro de memória $d \in \mathbb{R}$, tal que a diferenciação de ordem $K > d - 0.5$, denotada por $\Delta^K y_t$, é estacionária com função de densidade espectral dada por

$$f_{\Delta^K y_t}(\lambda) = |1 - e^{-i\lambda}|^{2(K-d)} f^*(\lambda), \quad \lambda \in (-\pi, \pi),$$

onde $f^*(\lambda)$ é uma função não-negativa, par, contínua, integrável em $(-\pi, \pi)$ e tal que $f^*(0) > 0$. O símbolo Δy_t representa a diferenciação de primeira ordem dada por $\Delta y_t = y_t - y_{t-1}$, assim a diferenciação de ordem K pode ser obtida recursivamente. Essa definição é sugerida por Hurvich & Ray (1995), uma definição alternativa tem sido proposta por Robinson (1994), Tanaka (1999), Shimotsu & Phillips (2005) entre outros.

O parâmetro de memória d caracteriza o comportamento do espectro nas frequências próximas de zero, portanto quando d assume valores diferentes de zero provoca mudanças significativas na estrutura de dependência de um processo; por exemplo o processo $\{y_t; t = 1, 2, \dots, n\}$ é estacionário e invertível quando $d \in (-1.0, 0.5)$ e

$$f_{y_t}(\lambda) = |1 - e^{-i\lambda}|^{-2d} f^*(\lambda), \quad \lambda \in (-\pi, \pi), \quad (1)$$

onde $f^*(\lambda) = \frac{\sigma^2}{2\pi} \left| \frac{\Theta(e^{i\lambda})}{\Phi(e^{i\lambda})} \right|^2$ e as raízes dos polinômios $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p = 0$ e $\Theta(z) = 1 - \theta_1 z - \dots - \theta_p z^q = 0$ são não-comuns e estão fora do círculo unitário. Quando $d \geq 0.5$ o processo se diz não-estacionário e a função $f_{y_t}(\lambda)$ não é integrável. Em ambos os casos, a função de densidade espectral do processo $\{y_t; t = 1, 2, \dots, n\}$ satisfaz $f_{y_t}(\lambda) \sim |\lambda|^{-2d} f^*(0)$ quando $\lambda \rightarrow 0^+$. O símbolo “ \sim ” significa que o quociente entre o lado esquerdo e o lado direito tende a 1 quando $\lambda \rightarrow 0$.

O processo $\{y_t; t = 1, 2, \dots, n\}$ com tais características é conhecido na literatura como ARFIMA e é utilizado na modelagem de séries que possuem propriedade de memória longa (ver, e.g. Granger & Joyeux (1980) e Hosking (1981)). A propriedade de memória longa ocorre em séries que apresentam correlações estatisticamente significativas mesmo para observações distantes; equivalentemente, o espectro apresenta singularidade para frequências próximas de 0. Neste documento são considerados métodos de estimação do parâmetro d em processos com representação ARFIMA gaussianos.

Existem vários estimadores do parâmetro de diferenciação fracionária d propostos na literatura que podem ser classificados em paramétricos e semi-paramétricos. Os primeiros envolvem a estimação simultânea dos parâmetros do modelo, em geral utilizando o método de máxima verossimilhança; ver, e.g., Fox & Taqqu (1986), entre outros. Nos procedimentos semi-paramétricos, a estimação dos parâmetros do modelo é realizada em dois passos: primeiro estima-se o parâmetro de memória longa d e, posteriormente, estimam-se os parâmetros autorregressivos e de médias móveis. O estimador mais popular dessa classe é o estimador proposto por Geweke & Porter-Hudak (1983); variantes foram desenvolvidas por Chen et al. (1994), Reisen (1994), Robinson (1995a, 1995b), entre outros.

2.1. Estimador Log-periodograma (LP). Seja $f(\lambda_j)$ a função definida na Eq. 1 para $\lambda_j = \frac{2\pi j}{n}$, $j = 1, \dots, \lfloor \frac{n}{2} \rfloor$, onde n é o tamanho amostral e $\lfloor \cdot \rfloor$ denota a função parte inteira. Sejam $f(\lambda_j) := f_j$ e $f^*(\lambda_j) := f_j^*$. O logaritmo de f_j pode ser escrito como:

$$\ln f_j = \ln f^*(0) - d \ln |1 - e^{-i\lambda_j}|^2 + \ln \frac{f_j^*}{f^*(0)}. \quad (2)$$

Seja $I_j := I(\lambda_j) = |w(\lambda_j)|^2$, onde $w(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n y_t e^{i\lambda_j t}$ representa a transformada discreta de Fourier (TDF). A função $I(\lambda_j)$ é chamada *periodograma* e representa um estimador assintoticamente não-viesado para a função de densidade espectral f_j . As propriedades do periodograma clássico têm sido amplamente estudadas em processos estacionários com propriedade de memória curta e longa (ver e.g. Hannan (1973), Yajima (1989), Hurvich & Beltrão (1993) e Robinson (1995b)).

Adicionando $\ln I_j = \ln \frac{I_j}{f_j} + \ln f_j$ na Eq. 2, obtém-se a equação:

$$\ln I_j = \ln f^*(0) - d \ln |1 - e^{-i\lambda_j}|^2 + \ln \frac{I_j |1 - e^{-i\lambda_j}|^{2d}}{f^*(0)}, \quad (3)$$

que sugere a equação de regressão dada por

$$\ln I_j = \beta_0 + \beta_1 \ln |1 - e^{-i\lambda_j}|^2 + e_j, \quad j = 1, 2, \dots, g(n),$$

onde $\beta_0 = \ln f^*(0)$ e $\beta_1 = -d$. Note que, para frequências próximas de zero e $g(n) = o(n)$, então

$$f_j \sim |\lambda_j|^{-2d} f^*(0), \quad (4)$$

assim, $e_j \sim \ln \frac{I_j}{f_j}$, para $j = 1, 2, \dots, g(n)$.

Geweke & Porter-Hudak (1983) sugerem um estimador semiparamétrico para d , dado por

$$d_{LP} = -\frac{\sum_{i=1}^{g(n)} (v_i - \bar{v}) \ln I_i}{\sum_{i=1}^{g(n)} (v_i - \bar{v})^2}, \quad (5)$$

onde $v_j = \ln |1 - e^{-i\lambda_j}|^2$, $\bar{v} = \frac{1}{g(n)} \sum v_j$ e $g(n)$ é chamado de *bandwidth* e corresponde ao número de frequências utilizadas na regressão.

Hurvich et al. (1998), sob algumas condições de regularidade, calculam um valor ótimo do *bandwidth* tal que $g(n) = O(n^{4/5})$. As propriedades assintóticas do estimador LP foram derivadas por Robinson (1995b) e Hurvich et al. (1998), para o caso estacionário. No contexto não-estacionário e considerando a relação $f_{y_t}(\lambda_j) \sim |\lambda_j|^{-2d} f^*(0)$ quando $\lambda \rightarrow 0^+$, Velasco (1999b) estende os resultados obtidos por Robinson (1995b) e mostra a consistência do estimador LP para $d \in (0.5, 1]$. Kim & Phillips (2006) mostram que para valores $d > 1$ o estimador LP converge em probabilidade para 1. Phillips (1999) prova a normalidade assintótica do estimador para $d \in (0.5, 1)$, i.e.

$$\sqrt{g(n)}(d_{LP} - d) \xrightarrow{\mathcal{D}} N\left(0, \frac{\pi^2}{24}\right),$$

onde $\xrightarrow{\mathcal{D}}$ denota convergência em distribuição. No caso da presença de raiz unitária, Phillips (2007) mostra que o estimador LP assintoticamente apresenta distribuição normal mista com $\text{var}(d_{LP}) = 0.3948g^{-1}(n)$.

3. ESTIMADORES ROBUSTOS

Seja $\{z_t; t = 1, 2, \dots, n\}$ um processo contaminado por dados atípicos de tipo aditivo com representação dada por

$$z_t = y_t + v_t, \quad (6)$$

onde a sequência $\{v_t; t = 1, 2, \dots, n\}$ é tal que v_t segue uma distribuição de misturas de normais, dada por $F_v = (1-\gamma)\delta_0 + \gamma\mathcal{N}(0, \sigma_v^2)$, onde δ_0 representa a distribuição de massa pontual concentrada em zero e $\sigma_v > 0$. O processo não-contaminado $\{y_t; t = 1, 2, \dots, n\}$ segue uma representação ARFIMA gaussiano.

A Proposição 2 em Fajardo et al. (2009) mostra que a presença de dados atípicos em processos estacionários provoca aumento significativo no viés do estimador do espectro em função da magnitude dos dados atípicos. As estimativas dos parâmetros obtidas através dos métodos de estimação baseados no periodograma são seriamente afetadas. Quando o processo $\{y_t; t = 1, 2, \dots, n\}$ é não-estacionário obtém-se um resultado equivalente:

Teorema 1. *Sejam z_1, z_2, \dots, z_n variáveis aleatórias geradas do modelo definido na Eq. 6. O periodograma avaliado nas frequências $\lambda_j = \frac{2\pi j}{n}$, com $0 < j < \lfloor \frac{n}{2} \rfloor$, é dado por*

$$I_{zj} = I_{yj} + I_{vj} + 2\Re\{w_{yj}\overline{w_{vj}}\}, \quad (7)$$

onde $I_y(\lambda_j) := I_{yj}$ é dado por

$$I_{yj} = \left|1 - e^{i\lambda_j}\right|^{-2} \left[I_{xj} + \frac{1}{2\pi n} y_n^2 - 2\Re\left\{w_{xj} \frac{e^{-i\lambda_j}}{\sqrt{2\pi n}} y_n\right\} \right], \quad (8)$$

$I_{xj} = |w_{xj}|^2$ com $w_{xj} = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n x_t e^{i\lambda_j t}$ a transformada discreta de Fourier (TDF) de $\{x_t\}$ e $I_{vj} = |w_{vj}|^2$ com $w_{vj} = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n v_t e^{i\lambda_j t}$. $\Re\{\cdot\}$ representa a parte real de um número complexo.

O resultado dado na Eq. 7 mostra que o periodograma de séries contaminadas apresenta um termo adicional em função da magnitude das observações atípicas e da última observação y_n . A influência desses termos afeta significativamente o viés do estimador do espectro e consequentemente as estimativas dos parâmetros dos modelos ajustados. Na prática, a tendência do estimador LP a subestimar o verdadeiro valor do parâmetro de memória se explica pela presença do termo em função de σ_v^2 . Para reduzir o efeito dos dados atípicos na estimação dos parâmetros do modelo, sugere-se a utilização de estimadores robustos para o espectro do processo. Como sugerido por Fajardo et al. (2009) para o caso estacionário, o uso de um estimador robusto do espectro no estimador LP, definido na Eq. 5, pode garantir estimativas com menor viés para o parâmetro de memória em séries contaminadas. A seguir são apresentados estimadores robustos para o parâmetro de memória de processos contaminados por *outliers* de tipo aditivo.

3.1. Estimador tipo M do parâmetro de memória. Seguindo a metodologia sugerida por Geweke & Porter-Hudak (1983) sugere-se a equação de regressão dada por

$$\ln \tilde{I}_j = \beta_0 + \beta_1 \ln |1 - e^{-i\lambda_j}|^2 + e_j, \quad j = 1, 2, \dots, g(n),$$

onde $\beta_0 = \ln f^*(0)$ e $\beta_1 = -d$. Note que, para frequências próximas de zero e $g(n) = o(n)$, então

$$f_j \sim |\lambda_j|^{-2d} f^*(0),$$

assim, $e_j \sim \ln \frac{\tilde{I}_j}{f_j}$, para $j = 1, 2, \dots, g(n)$. A definição do estimador robusto para o parâmetro de diferenciação fracionária é dado por

$$d_M = -\frac{\sum_{i=1}^{g(n)} (v_i - \bar{v}) \ln \tilde{I}_i}{\sum_{i=1}^{g(n)} (v_i - \bar{v})^2}, \quad (9)$$

onde $v_j = \ln |1 - e^{-i\lambda_j}|^2$, $\bar{v} = \frac{1}{g(n)} \sum v_j$ e $g(n)$ como definido na Eq. 5. O estimador \tilde{I}_i representa o *periodograma tipo M* definido como

$$\tilde{I}_j = \frac{1}{2\pi} I_j^*, \quad (10)$$

onde $I_j^* = \frac{n}{4} \{T_{n0}^2 + T_{n1}^2\}$, $\lambda_j = \frac{2\pi j}{n}$ com $j = 1, 2, \dots, \lfloor \frac{n}{2} \rfloor$. Para cada j , o vetor $T_n(F_n; \lambda_j) := T_{nj} = (T_{nj,0}, T_{nj,1})'$ é um M-estimador para $\theta = (\theta_{0j}, \theta_{1j})'$ que satisfaz

$$\frac{1}{n} \sum_{t=1}^n \rho \left(\frac{z_t - \mathbf{x}_t' T_{nj}}{\hat{\sigma}} \right) = \min_{\theta \in \mathbb{R}^2} \frac{1}{n} \sum_{t=1}^n \rho \left(\frac{z_t - \mathbf{x}_t' \theta}{\hat{\sigma}} \right), \quad (11)$$

onde $\mathbf{x}_t = (\cos(\lambda_j t), \sin(\lambda_j t))'$, F_n representa a função de distribuição empírica e $\hat{\sigma}$ um estimador robusto para $\sigma = \sqrt{\text{var}(z_t - \mathbf{x}_t' T_{nj})}$. Estimativas robustas para σ podem ser encontradas em Rousseeuw & Croux (1993). A função $\rho : \mathbb{R}^2 \rightarrow \mathbb{R}$ é chamada de *função de perda* e pode ser qualquer função convexa que satisfaz certas condições para diferenciabilidade e permita obter estimativas robustas para o vetor θ . Huber (1964) mostra que a convexidade da função ρ permite provar a consistência e normalidade assintótica do estimador T_{nj} sob condições mínimas, assim como unicidade e existência da solução da Eq. 11. A definição de M-estimador na Eq. 11 deve-se a Huber (1964), e diferentes aplicações dos mesmos têm sido desenvolvidas na literatura, várias referências podem ser encontradas nas publicações de Hampel et al. (1986), Huber (2004), Maronna et al. (2006) e Bai et al. (1997) entre outros.

As propriedades assintóticas do estimador definido na Eq. 10 podem ser mostradas sob algumas condições de regularidade, quando o processo $\{y_t; t = 1, 2, \dots, n\}$ é estacionário. As condições de regularidade são apresentadas por Arcones (1998) e Bai et al. (1992) e permitem a abrangência de uma classe ampla de funções de perda. No caso não-estacionário, Wu (2007) sugere uma condição adicional por causa da dependência entre os erros na representação espectral do processo. A suposição assume que a dependência entre erros é de curto prazo e que a contribuição acumulada

de uma variável de erro é finita quando são calculadas previsões futuras. A condição dada por Wu (2007) permite provar a normalidade assintótica e consistência do vetor T_n para o caso não-estacionário (ver Wu (2007), Teorema 1).

As condições de regularidade garantem a existência de uma única solução para a Eq. 11 e permitem mostrar que a estatística T_{nj} também é solução do sistema de equações dado por

$$\sum_{t=1}^n \varrho \left(\frac{y_t - \mathbf{x}'_t \theta}{\hat{\sigma}} \right) \mathbf{x}_t = 0, \quad (12)$$

ou seja, o M-estimador T_{nj} pode ser encontrado como solução das Eqs. 11 ou 12 de forma equivalente. A função $\varrho(\cdot)$ definida na Eq. 12 permite considerar uma classe ampla de estimadores do parâmetro d , por exemplo se $\varrho(x) = x$ obtém-se o estimador LP definido na Eq. 5. Neste trabalho, para o cálculo do periodograma, definido na Eq. 10, são consideradas as seguintes funções:

1. Periodograma bisquare: obtido a partir de

$$\varrho(x) = 0.2734x [1 - 0.0456x^2]^2 \text{ se } |x| \leq 4.6850.$$

2. Periodograma Hampel: obtido a partir de

$$\varrho(x) = \begin{cases} x, & \text{se } 0 \leq |x| \leq 2; \\ 2 \operatorname{sgn}(x), & \text{se } 2 \leq |x| \leq 4; \\ \frac{8-|x|}{2} \operatorname{sgn}(x), & \text{se } 4 \leq |x| \leq 8; \\ 0, & \text{se } 8 \leq |x| \end{cases}$$

3. Periodograma Huber: obtido por

$$\varrho(x) = \begin{cases} x, & \text{if } |x| \leq 1.345; \\ 1.345 \operatorname{sgn}(x), & \text{if } |x| > 1.345. \end{cases}$$

Os estimadores do parâmetro d serão denotados por LS quando utilizado o periodograma clássico, BS quando utilizado o periodograma bisquare, HP quando utilizado o periodograma Hampel e HU quando utilizado o periodograma Huber.

3.2. Estimador tipo L para o parâmetro d . Fajardo et al. (2009) sugerem uma variante robusta do estimador LP dada por

$$d_L = - \frac{\sum_{j=1}^{g(n)} (v_j - \bar{v}) \log I_Q(\lambda_j)}{\sum_{j=1}^{g(n)} (v_j - \bar{v})^2}, \quad (13)$$

onde $v_j = \log \left\{ 2 \sin \left(\frac{\lambda_j}{2} \right) \right\}^2$, $\bar{v} = \frac{1}{g(n)} \sum v_j$ e $I_Q(\lambda_j) = \frac{1}{2\pi} \sum_{|s| \leq r} \kappa \left(\frac{h}{r} \right) \widehat{R}_Q(s) e^{ih\lambda_j}$. A função $\kappa(\cdot)$ é chamada de *janela* de alisamento e $r = n^\beta$, $0 < \beta < 1$. Os autores sugerem a aplicação da *janela truncada* dada por $\kappa(x) = 1$ se $|x| \leq r$ com $r = n^{0.7}$.

A função $\widetilde{R}(\cdot)$ representa um estimador robusto para a autocovariância dado por

$$\widetilde{R}(h) = \frac{1}{4} [Q_{n-h}^2(\mathbf{u} + \mathbf{v}) - Q_{n-h}^2(\mathbf{u} - \mathbf{v})],$$

onde \mathbf{u} e \mathbf{v} são vetores que constam das primeiras e últimas $n - h$ observações, respectivamente. O estimador $Q_n(\cdot)$ calcula a k -ésima estatística de ordem das $\binom{n}{2}$ distâncias $\{|z_i - z_j|, i < j\}$, i.e. $Q_n(z) = c_\Phi \cdot \{|z_i - z_j|; i < j\}_{(k)}$, onde $z = (z_1, z_2, \dots, z_n)'$ representa o vetor de dados, c_Φ é uma constante para garantir a propriedade de consistência ($c_\Phi = 2.2191$ para distribuição normal) e $k = \left\lfloor \frac{\binom{n}{2} + 2}{4} \right\rfloor + 1$.

Na análise de séries temporais, a aplicação de *janelas* no estimador do espectro é um procedimento comumente utilizado para obter estimadores consistentes da densidade espectral de processos

estacionários. Fajardo et al. (2009) sugerem a aplicação da *janela truncada* para redução do efeito irregular do estimador robusto da função de autocovariância para as últimas defasagens. Fajardo (2011) sugere a aplicação de *janelas* de alisamento alternativas para melhorar o desempenho do estimador robusto do espectro, as quais são apresentadas a seguir:

i. janela de Parzen:

$$\kappa(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & \text{se } |x| \leq 0.5, \\ 2(1 - |x|)^3, & \text{se } 0.5 < |x| \leq 1.0, \\ 0, & \text{se } |x| > 1. \end{cases}$$

ii. janela de Tuckey-Hamming (TH): $\kappa(x) = 0.54 + 0.46 \cos(\pi x)$, $|x| \leq 1$.

iii. janela de Bartlett: $\kappa(x) = 1 - |x|$, $|x| \leq 1$.

Ao selecionar a *janela* alisadora deve-se levar em conta a escolha do valor do r . Na prática, sugere-se a utilização de critérios estatísticos adequados para seleção ótima de tal valor. Para seleção do valor β para cada *janela* acima definida, ver análise detalhada em Fajardo (2011).

4. EXPERIMENTOS DE MONTE CARLO

O comportamento empírico do estimador robusto é analisado através de simulações de Monte Carlo. As séries observadas foram geradas de um processo $\{z_t; t = 1, 2, \dots, n\}$ com representação dada na Eq. 6 com $\sigma_v = 10$ e valores $\gamma = 0, 0.05, 0.1, 0.25$. O processo não-contaminado $\{y_t; t = 1, 2, \dots, n\}$ é gerado a partir de um processo ARFIMA(p, d, q) gaussiano dado por

$$y_t = (1 - B)^{-d} u_t, \quad t = 1, 2, \dots, n, \quad (14)$$

onde o filtro de diferenciação fracionária $(1 - B)^d$, para $d \in \mathbb{R}$, é definido pela expansão binomial $(1 - B)^d = \sum_{j=0}^{\infty} \pi_j B^j$, onde $\pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}$ e $\Gamma(\cdot)$ é a função gamma. O processo $\{u_t; t = 1, 2, \dots, n\}$ com representação ARMA(p, q) apresenta ruído branco gaussiano com média 0 e variância 1. Parâmetros $p = q = 1$ com valores $\phi = 0.2$ e $\theta = 0.5$ para as componente autorregressiva e de médias móveis, respectivamente. Para a geração de dados com representação dada na Eq. 14, segue-se a metodologia sugerida por Hosking (1981) para valores $d = 0.1, 0.3, 0.45, 0.5, 0.6, 0.8, 1.0$ e 1.5. Foram obtidas 10000 realizações com tamanhos de amostra $n = 300$ e 800 para o processo contaminado com outliers de tipo aditivo, definido na Eq. 6. As simulações foram realizadas na linguagem de programação matricial `0x` (ver <http://www.doornik.com>).

Os estimadores do parâmetro de memória calculados para cada realização do processo foram obtidos a partir do estimador definido na Eq. 9 com $g(n) = n^\alpha$, $\alpha = 0.5$. A escolha de um valor ótimo para o *bandwidth* na presença de contaminação nos dados ainda é um problema aberto. Neste documento optou-se por utilizar o valor $\alpha = 0.5$ para diminuir o viés do estimador LP no caso contaminado, como sugerem os resultados empíricos obtidos por Haldrup & Nielsen (2007).

4.1. Estimadores tipo M. Na Tabela 1 são apresentadas as estimativas obtidas do parâmetro d quando o processo sob análise é estacionário. A tabela mostra a média, o viés e o erro quadrático médio (EQM) dos estimadores *LS*, *BS*, *HP* e *HU* para as realizações do processo com diferentes níveis de contaminação. Neste cenário se evidencia o baixo desempenho do periodograma clássico sob contaminação nos dados, mesmo para menores níveis de contaminação. Como mostrado por Fajardo et al. (2009), no caso de amostra finita o estimador LS é afetado significativamente na presença de pelo menos uma observação atípica nos dados. Os resultados mostrados na tabela indicam um aumento superior a 30% no EQM do estimador LS quando o conjunto de observações apresenta contaminação.

O desempenho dos estimadores robustos na ausência de dados atípicos indica que podem ser considerados alternativas razoáveis para estimação do parâmetro d quando o processo sob análise não apresenta contaminação. Em processos contaminados destaca-se o desempenho do estimador

		$d = 0.1$			$d = 0.3$			$d = 0.45$		
Est	n	média	viés	EQM	média	viés	EQM	média	viés	EQM
<i>LS</i>	300	0.0896	0.0104	0.0364	0.2949	0.0051	0.0376	0.4587	-0.0087	0.0369
<i>BS</i>		0.0905	0.0095	0.0352	0.2945	0.0055	0.0357	0.4547	-0.0047	0.0375
<i>HP</i>		0.0889	0.0111	0.0369	0.2947	0.0053	0.0358	0.4590	-0.0090	0.0373
<i>HU</i>		0.0911	0.0089	0.0358	0.2944	0.0056	0.0354	0.4539	-0.0039	0.0372
<i>LS</i>	800	0.1067	-0.0067	0.0179	0.3103	-0.0103	0.0189	0.4551	-0.0051	0.0221
<i>BS</i>		0.1021	-0.0021	0.0183	0.3070	-0.0070	0.0194	0.4506	-0.0006	0.0234
<i>HP</i>		0.1053	-0.0053	0.0177	0.3094	-0.0094	0.0190	0.4541	-0.0041	0.0225
<i>HU</i>		0.1023	-0.0023	0.0184	0.2919	0.0081	0.0200	0.4506	-0.0006	0.0236
Contaminação 5%										
<i>LS</i>	300	0.0213	0.0787	0.0479	0.1495	0.1505	0.0627	0.3048	0.1452	0.0622
<i>BS</i>		0.0936	0.0064	0.0362	0.2864	0.0136	0.0369	0.4488	0.0012	0.0402
<i>HP</i>		0.0918	0.0082	0.0357	0.2855	0.0145	0.0367	0.4451	0.0049	0.0406
<i>HU</i>		0.0854	0.0146	0.0366	0.2782	0.0218	0.0343	0.4417	0.0083	0.0383
<i>LS</i>	800	0.0255	0.0745	0.0254	0.1582	0.1418	0.0409	0.3259	0.1241	0.0369
<i>BS</i>		0.0967	0.0033	0.0190	0.3001	-0.0001	0.0189	0.4416	0.0084	0.0234
<i>HP</i>		0.0944	0.0056	0.0201	0.2992	0.0008	0.0184	0.4428	0.0072	0.0216
<i>HU</i>		0.0876	0.0124	0.0230	0.2919	0.0081	0.0200	0.4361	0.0139	0.0228
Contaminação 10%										
<i>LS</i>	300	0.0016	0.0984	0.0532	0.0832	0.2168	0.0897	0.2167	0.2333	0.1013
<i>BS</i>		0.0908	0.0092	0.0392	0.2815	0.0185	0.0395	0.4190	0.0309	0.0416
<i>HP</i>		0.0834	0.0166	0.0434	0.2698	0.0302	0.0423	0.4081	0.0419	0.0444
<i>HU</i>		0.0659	0.0341	0.0420	0.2512	0.0488	0.0440	0.4003	0.0497	0.0439
<i>LS</i>	800	0.0143	0.0857	0.0292	0.1139	0.1861	0.0565	0.2852	0.1648	0.0483
<i>BS</i>		0.0948	0.0052	0.0196	0.2959	0.0041	0.0187	0.4418	0.0082	0.0227
<i>HP</i>		0.0918	0.0082	0.0200	0.2881	0.0119	0.0208	0.4384	0.0116	0.0225
<i>HU</i>		0.0868	0.0132	0.0204	0.2808	0.0192	0.0202	0.4331	0.0169	0.0222
Contaminação 25%										
<i>LS</i>	300	0.0117	0.0883	0.0441	0.0489	0.2511	0.0988	0.1250	0.3249	0.1432
<i>BS</i>		0.0911	0.0089	0.0407	0.2568	0.0432	0.0414	0.3683	0.0817	0.0473
<i>HP</i>		0.0691	0.0309	0.0442	0.2252	0.0748	0.0457	0.3188	0.1312	0.0609
<i>HU</i>		0.0658	0.0342	0.0405	0.2182	0.0818	0.0442	0.3314	0.1186	0.0559
<i>LS</i>	800	0.0042	0.0958	0.0322	0.0566	0.2434	0.0812	0.1749	0.2751	0.0974
<i>BS</i>		0.0560	0.0440	0.0231	0.2434	0.0566	0.0236	0.3833	0.0667	0.0264
<i>HP</i>		0.0454	0.0546	0.0236	0.2136	0.0864	0.0259	0.3455	0.1045	0.0321
<i>HU</i>		0.0456	0.0544	0.0240	0.2120	0.0880	0.0283	0.3547	0.0953	0.0308

TABELA 1. Estimadores tipo M: Estimativas do parâmetro d no modelo ARFIMA(0, d , 0) estacionário.

BS por apresentar menor viés e menor EQM entre os estimadores robustos propostos, mesmo para altas porcentagens de contaminação. O aumento no nível de contaminação permite observar o aumento no viés do estimador *LS* sugerindo que o uso do mesmo deve ser evitado nas análises práticas.

Na região de não-estacionariedade o comportamento dos estimadores é apresentado nas Tabelas 2 e 3. Os resultados evidenciam a superioridade dos estimadores robustos na estimação do parâmetro de memória quando as observações são contaminadas por dados atípicos.

De forma análoga ao caso estacionário, o estimador *BS* se destaca por apresentar menor viés mesmo para altos níveis de contaminação. Os estimadores *HP* e *HU* apresentam-se como estimadores alternativos para d na presença de dados atípicos nas observações. Em dados contaminados o uso do estimador *LS* não é conveniente e sugere-se o uso das alternativas robustas para aplicações práticas. A Tabela 3 mostra os resultados obtidos para $d = 1$. Em séries com raiz unitária os estimadores do parâmetro de memória apresentam menores valores do EQM em comparação com os valores obtidos para $d \in [0.5, 1)$, isto evidencia uma equivalência com os resultados apresentados por Phillips (2007) para o caso do estimador LP. Para valores $d > 1$ as estimativas obtidas a partir dos estimadores sob estudo tendem a 1, confirmando os resultados obtidos por Hurvich & Ray (1995) e Kim & Phillips (2006) para séries sem contaminação. No caso contaminado, o aumento na variância do processo quando o valor do $d = 1$, provoca a diminuição significativa do efeito causado pela presença de dados atípicos nas séries e faz com que as estimativas obtidas a partir dos estimadores clássico e robustos sejam similares.

A Figura 1 apresenta o boxplot das estimativas obtidas para as séries observadas com diferentes valores de d e diferentes níveis de contaminação. As porcentagens de contaminação são representados

		$d = 0.5$			$d = 0.6$			$d = 0.8$		
Est	n	média	viés	EQM	média	viés	EQM	média	viés	EQM
<i>LS</i>	300	0.5136	-0.0136	0.0414	0.6248	-0.0248	0.0403	0.8301	-0.0301	0.0400
<i>BS</i>		0.5004	-0.0004	0.0421	0.6188	-0.0188	0.0414	0.8169	-0.0169	0.0416
<i>HP</i>		0.5097	-0.0097	0.0410	0.6245	-0.0245	0.0406	0.8255	-0.0255	0.0390
<i>HU</i>		0.5008	-0.0008	0.0425	0.6180	-0.0180	0.0414	0.8222	-0.0222	0.0380
<i>LS</i>	800	0.5062	-0.0062	0.0231	0.6182	-0.0182	0.0219	0.8262	-0.0262	0.0234
<i>BS</i>		0.4986	0.0014	0.0230	0.6108	-0.0108	0.0218	0.8156	-0.0156	0.0242
<i>HP</i>		0.5057	-0.0057	0.0228	0.6154	-0.0154	0.0213	0.8233	-0.0233	0.0223
<i>HU</i>		0.4980	0.0020	0.0236	0.6103	-0.0103	0.0220	0.8173	-0.0173	0.0221
Contaminação 5%										
<i>LS</i>	300	0.3466	0.1534	0.0668	0.4807	0.1193	0.0574	0.7293	0.0707	0.0483
<i>BS</i>		0.4828	0.0172	0.0438	0.5843	0.0157	0.0413	0.7728	0.0272	0.0378
<i>HP</i>		0.4828	0.0172	0.0412	0.5803	0.0197	0.0398	0.7695	0.0305	0.0374
<i>HU</i>		0.4776	0.0224	0.0438	0.5782	0.0218	0.0406	0.7737	0.0263	0.0394
<i>LS</i>	800	0.3917	0.1083	0.0353	0.5255	0.0745	0.0286	0.7749	0.0251	0.0245
<i>BS</i>		0.4882	0.0118	0.0240	0.5935	0.0065	0.0224	0.7904	0.0096	0.0228
<i>HP</i>		0.4902	0.0098	0.0231	0.5943	0.0057	0.0225	0.7937	0.0063	0.0224
<i>HU</i>		0.4852	0.0148	0.0239	0.5898	0.0102	0.0221	0.7870	0.0130	0.0224
Contaminação 10%										
<i>LS</i>	300	0.2744	0.2256	0.0885	0.3855	0.2145	0.0926	0.6701	0.1299	0.0562
<i>BS</i>		0.4745	0.0255	0.0395	0.5502	0.0498	0.0467	0.7378	0.0622	0.0397
<i>HP</i>		0.4584	0.0416	0.0371	0.5308	0.0692	0.0447	0.7332	0.0668	0.0406
<i>HU</i>		0.4658	0.0342	0.0375	0.5365	0.0635	0.0458	0.7421	0.0579	0.0388
<i>LS</i>	800	0.3146	0.1854	0.0565	0.4759	0.1241	0.0378	0.7322	0.0678	0.0251
<i>BS</i>		0.4819	0.0181	0.0215	0.5804	0.0196	0.0222	0.7649	0.0351	0.0206
<i>HP</i>		0.4721	0.0279	0.0209	0.5706	0.0294	0.0226	0.7598	0.0402	0.0203
<i>HU</i>		0.4673	0.0327	0.0223	0.5754	0.0246	0.0226	0.7643	0.0357	0.0203
Contaminação 25%										
<i>LS</i>	300	0.1543	0.3457	0.1627	0.2640	0.3359	0.1516	0.5182	0.2818	0.1247
<i>BS</i>		0.4038	0.0962	0.0433	0.4729	0.1271	0.0587	0.6262	0.1738	0.0674
<i>HP</i>		0.3639	0.1361	0.0545	0.4156	0.1844	0.0758	0.5925	0.2075	0.0846
<i>HU</i>		0.3708	0.1292	0.0537	0.4516	0.1484	0.0638	0.6251	0.1749	0.0681
<i>LS</i>	800	0.2361	0.2639	0.0896	0.3496	0.2504	0.0859	0.6496	0.1504	0.0464
<i>BS</i>		0.4392	0.0608	0.0215	0.5021	0.0979	0.0315	0.6992	0.1009	0.0317
<i>HP</i>		0.4003	0.0997	0.0268	0.4662	0.1338	0.0387	0.6860	0.1140	0.0345
<i>HU</i>		0.4146	0.0854	0.0238	0.4902	0.1098	0.0340	0.7030	0.0970	0.0312

TABELA 2. Estimadores tipo M: Estimativas do parâmetro d no modelo ARFIMA(0, d , 0) não-estacionário.

pelos cores: branca para 0%, cinza escuro para 5%, cinza médio para 10% e cinza claro para 25% de contaminação. A figura evidencia o aumento significativo no viés do estimador *LS* e destaca-se o desempenho dos estimadores robustos enquanto a viés e EQM tanto para séries contaminadas quanto para séries sem contaminação. Nos resultados não são encontradas diferenças significativas no comportamento dos estimadores robustos aqui utilizados. Porém, na prática sugere-se o uso do estimador *BS* por apresentar menor custo computacional.

As estimativas do estimador tipo *M* quando o modelo ARFIMA envolve componentes autorregressivas e de médias móveis apresentam-se na Tabela 4. Os resultados obtidos mostram que o estimador apresenta um leve aumento no viés em séries sem contaminação. Em séries contaminadas, as estimativas obtidas através do estimador *LS* são significativamente influenciadas pela presença de dados atípicos ao longo da série. Em aplicações práticas o uso do estimador tipo *M* é recomendável pelo seu desempenho em amostras finitas, mesmo para níveis de contaminação de 10%. Como observado por Fajardo et al. (2009), para valores de ϕ próximos de 1, o estimador *LS* evidencia um aumento significativo no viés que ocasiona uma tendência a superestimar o verdadeiro valor do parâmetro de memória. Em modelos ARFIMA(p , d , q) a superestimação devida a presença de componentes autorregressivas e a subestimação causada pelas observações atípicas na série, pode provocar o cálculo de estimativas espúreas do parâmetro d por parte dos estimadores tipo *M* e *L*. Como sugerido por Haldrup & Nielsen (2007) é recomendável o uso de valores reduzidos para o número de frequências na regressão que permite o cálculo dos estimadores semiparamétricos.

4.2. Estimadores tipo L. Fajardo et al. (2009) estudaram o desempenho do estimador na Eq. 13 na estimação do parâmetro de memória de processos ARFIMA estacionários. Os autores conduziram

		$d = 1.0$		
Est	n	média	viés	EQM
<i>LS</i>	300	1.0100	-0.0100	0.0337
<i>BS</i>		1.0021	-0.0021	0.0362
<i>HP</i>		1.0095	-0.0095	0.0336
<i>HU</i>		1.0037	-0.0037	0.0349
<i>LS</i>	800	1.0001	-0.0001	0.0162
<i>BS</i>		0.9924	0.0076	0.0174
<i>HP</i>		0.9995	0.0005	0.0157
<i>HU</i>		0.9921	0.0079	0.0169
Contaminação 5%				
<i>LS</i>	300	0.9593	0.0407	0.0380
<i>BS</i>		0.9584	0.0416	0.0374
<i>HP</i>		0.9630	0.0370	0.0365
<i>HU</i>		0.9615	0.0385	0.0369
<i>LS</i>	800	0.9814	0.0186	0.0180
<i>BS</i>		0.9746	0.0254	0.0197
<i>HP</i>		0.9815	0.0185	0.0173
<i>HU</i>		0.9778	0.0222	0.0198
Contaminação 10%				
<i>LS</i>	300	0.9032	0.0968	0.0451
<i>BS</i>		0.9106	0.0894	0.0437
<i>HP</i>		0.9089	0.0911	0.0430
<i>HU</i>		0.9122	0.0878	0.0426
<i>LS</i>	800	0.9573	0.0427	0.0201
<i>BS</i>		0.9585	0.0415	0.0210
<i>HP</i>		0.9593	0.0407	0.0202
<i>HU</i>		0.9597	0.0403	0.0203
Contaminação 25%				
<i>LS</i>	300	0.8110	0.1890	0.0785
<i>BS</i>		0.8244	0.1756	0.0704
<i>HP</i>		0.8203	0.1797	0.0744
<i>HU</i>		0.8279	0.1721	0.0687
<i>LS</i>	800	0.9141	0.0859	0.0260
<i>BS</i>		0.9191	0.0809	0.0256
<i>HP</i>		0.9191	0.0809	0.0249
<i>HU</i>		0.9202	0.0798	0.0246

TABELA 3. Estimadores tipo M: Estimativas do parâmetro d no modelo ARFIMA(0, d , 0) com raiz unitária.

Contaminação	0%			5%			10%			
	Estimador	média	viés	EQM	média	viés	EQM	média	viés	EQM
<i>LS</i>		0.2984	0.0016	0.0213	0.0929	0.2071	0.0673	0.0673	0.2327	0.0766
<i>BS</i>		0.2914	0.0086	0.0222	0.2808	0.0192	0.0209	0.2721	0.0279	0.0230
<i>HP</i>		0.2972	0.0028	0.0221	0.2786	0.0214	0.0209	0.2593	0.0407	0.0257
<i>HU</i>		0.2909	0.0091	0.0221	0.2686	0.0314	0.0228	0.2473	0.0527	0.0263

TABELA 4. Estimadores tipo M: Estimativas do parâmetro d no modelo ARFIMA(p , d , q) com $d = 0.3$, $\phi = 0.3$, $\theta = 0.5$ e $n = 800$.

estudos empíricos aplicando a *janela* truncada no estimador do espectro. A Tabela 5 apresenta as estimativas obtidas utilizando os estimadores robustos definidos usando as *janelas* definidas na Seção 3.2, para séries sem contaminação. Os estimadores são denotados por d_P , d_{TH} , d_B e d_{LPR} para

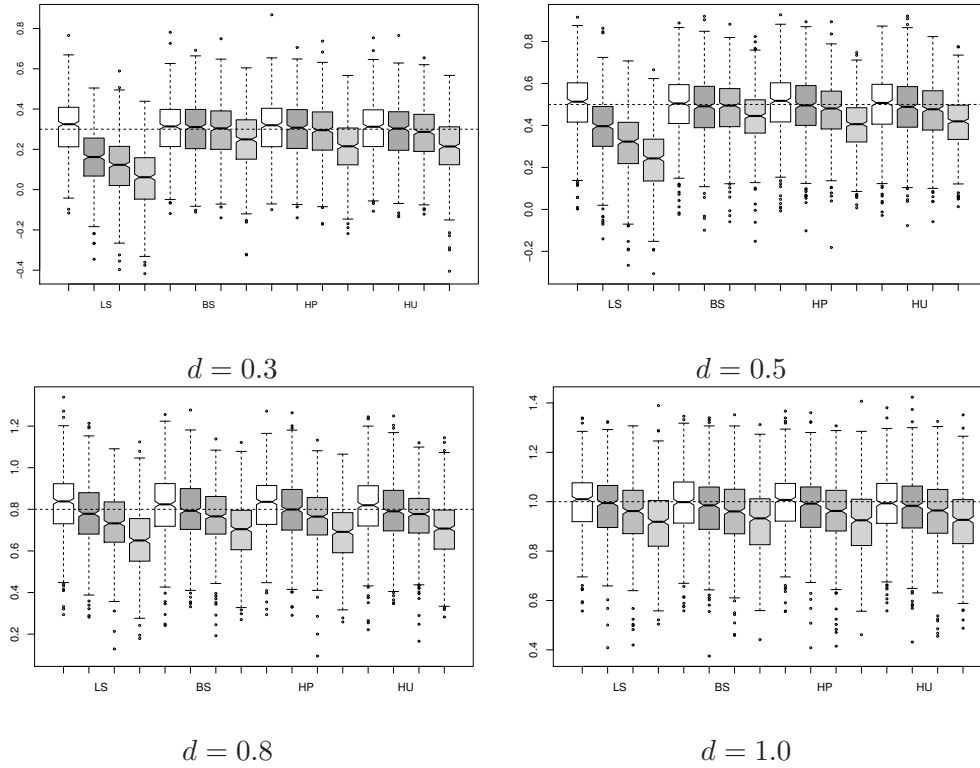


FIGURA 1. Média das estimativas do parâmetro d : 10000 realizações de tamanho 800 de um processo ARFIMA(0, d , 0) com diferentes níveis de contaminação (0% branco, 5% cinza escuro, 10% cinza médio, 25% cinza claro).

as *janelas* de Parzen, Tuckey-Hamming, Bartlett e truncada, respectivamente. Todas as *janelas* aplicadas com $\beta = 0.7$. Os resultados na tabela mostram que os estimadores robustos podem ser considerados alternativas razoáveis para estimação de d , mesmo em séries não contaminadas. Os resultados empíricos obtidos sugerem que o desempenho do estimador d_{LPR} , sugerido por Fajardo et al. (2009), pode ser melhorado com a aplicação de *janelas* alternativas.

Para o caso contaminado por dados atípicos de tipo aditivo (ver Tabela 6), os estimadores robustos se mostram resistentes ante a presença de *outliers* e evidenciam viés significativamente menor, se comparados com o estimador LS. A utilização de *janelas* diferentes à *janela truncada* permite evidenciar melhoras significativas na diminuição do viés do estimador d_L . Como mostram os resultados apresentados nas tabelas, pode-se sugerir o uso do periodograma suavizado com as *janelas* de Parzen e Bartlett, já que tendem a apresentar menor viés no caso de amostra finita.

Na região de não-estacionariedade a Tabela 7 apresenta as estimativas obtidas para $d = 0.5, 0.8, 1.0$ utilizando o estimador definido na Eq. 13 em séries com e sem contaminação. As colunas com d_{LPR} e d_{LPR_c} representam as estimativas obtidas através do estimador definido na Eq. 13, quando aplicada a *janela truncada* com $\beta = 0.7$ em séries com 0% e 5% de contaminação, respectivamente. Os resultados sugerem que o estimador robusto não deve ser usado em situações onde não é possível garantir a estacionariedade da série.

O comportamento empírico do estimador robusto do d evidencia a inconsistência do mesmo na região de não-estacionariedade. O baixo desempenho do estimador pode ser explicado porque $I_Q(\lambda)$ é função do estimador robusto da função de autocovariância $\hat{R}_Q(s)$, definida para processos estacionários (ver Ma & Genton (2000)). Destacam-se os resultados apresentados nas colunas à direita na Tabela 7, onde são apresentadas as estimativas do parâmetro de memória em séries com primeira diferença, denotadas por d_x . Observe-se que o estimador robusto sugerido por Fajardo et al. (2009)

Parâmetro	n		d_P	d_{TH}	d_B	d_{LPR}
$d = 0.1$	300	média	0.0796	0.0871	0.0828	0.0789
		viés	-0.0204	-0.0129	-0.0173	-0.0211
		EQM	0.0128	0.0126	0.0112	0.0115
	800	média	0.0933	0.0985	0.0957	0.0949
		viés	-0.0067	-0.0015	-0.0043	-0.0051
		EQM	0.0035	0.0048	0.0034	0.0037
$d = 0.3$	300	média	0.2880	0.2833	0.2857	0.2819
		viés	-0.0119	-0.0167	-0.0143	-0.0181
		EQM	0.0112	0.0110	0.0097	0.0124
	800	média	0.2985	0.2966	0.3001	0.2921
		viés	-0.0015	-0.0034	0.0001	-0.0079
		EQM	0.0031	0.0034	0.0031	0.0035
$d = 0.4$	300	média	0.4002	0.4049	0.4009	0.3932
		viés	0.0002	0.0049	0.0009	-0.0068
		EQM	0.0118	0.0118	0.0109	0.0119
	800	média	0.4037	0.4096	0.4066	0.4006
		viés	0.0037	0.0096	0.0066	0.0006
		EQM	0.0035	0.0040	0.0035	0.0042

TABELA 5. Estimadores tipo L: ARFIMA(0, d , 0) sem observações atípicas

Parâmetro	n		d_P	d_{TH}	d_B	d_{LPR}
$d = 0.1$	300	média	0.0874	0.0912	0.0922	0.0874
		viés	-0.0126	-0.0088	-0.0078	-0.0126
		EQM	0.0162	0.0144	0.0158	0.0161
	800	média	0.1115	0.1148	0.1127	0.1069
		viés	0.0115	0.0148	0.0127	0.0069
		EQM	0.0059	0.0059	0.0057	0.0059
$d = 0.3$	300	média	0.2806	0.2729	0.2796	0.2739
		viés	-0.0194	-0.0271	-0.0204	-0.0262
		d.p.	0.1028	0.0925	0.0964	0.1018
		EQM	0.0109	0.0093	0.0097	0.0111
	800	média	0.2934	0.2889	0.2928	0.2876
		viés	-0.0066	-0.0111	-0.0072	-0.0124
EQM		0.0034	0.0038	0.0031	0.0038	
$d = 0.4$	300	média	0.4331	0.4246	0.4339	0.4134
		viés	0.0331	0.0246	0.0339	0.0133
		EQM	0.0147	0.0136	0.0129	0.0121
	800	média	0.4414	0.4358	0.4409	0.4225
		viés	0.0413	0.0358	0.0409	0.0225
		EQM	0.0065	0.0057	0.0061	0.0057

TABELA 6. Estimadores tipo L: ARFIMA(0, d , 0) com observações atípicas.

é invariante para séries com primeira diferença. Lopes et al. (2004) através de resultados empíricos mostram que o estimador LP mantém a mesma propriedade para séries sem contaminação.

Aplicação: dados série IGP-DI Brasil. O IGP-DI representa o índice geral de preços disponibilidade interna, calculado pela Fundação Getulio Vargas. O período considerado para análise é agosto de 1994 – abril de 2011 (201 observações). A Figura 2 apresenta a dinâmica da série temporal com suas respectivas funções de autocorrelação (FAC), autocorrelação parcial (FACP) e o periodograma. As observações correspondentes a fevereiro de 1999 (4.44%), outubro de 2002 (4.21%) e novembro de 2002 (5.84%) podem causar aumento na variância dos dados, afetando as estimativas dos parâmetros do modelo ajustado. A FAC apresenta correlações estatisticamente

Parâmetro	n		d_{LPR}	d_{LPR_c}	Parâmetro	d_{LPR}	d_{LPR_c}
$d_y = 0.5$	300	média	0.4748	0.4659	$d_y = 0.5,$ $d_x = -0.5$	-0.3603	-0.4173
		viés	-0.0253	-0.0340		-0.1397	-0.0827
		EQM	0.0101	0.0096		0.0513	0.0684
	800	média	0.4927	0.4787		-0.3643	-0.3956
		viés	-0.0073	-0.0213		-0.1357	-0.1044
		MSE	0.0028	0.0032		0.0298	0.0406
$d_y = 0.8$	300	média	0.6448	0.6296	$d_y = 0.8,$ $d_x = -0.2$	-0.1906	-0.2211
		viés	-0.1552	-0.1704		-0.0094	0.0211
		EQM	0.0345	0.0382		0.0128	0.0206
	800	média	0.6316	0.6288		-0.2062	-0.2250
		viés	-0.1684	-0.1712		0.0062	0.0251
		EQM	0.0367	0.0366		0.0073	0.0123
$d_y = 1.0$	300	média	0.6659	0.6573	$d_y = 1.0,$ $d_x = 0.0$	-0.0273	-0.0426
		viés	-0.3341	-0.3427		0.0273	0.0426
		EQM	0.1276	0.1290		0.0127	0.0181
	800	média	0.6155	0.6129		-0.0107	-0.0222
		viés	-0.3845	-0.3871		0.0107	0.0222
		EQM	0.1657	0.1653		0.0041	0.0088

TABELA 7. Estimadores tipo L: Estimativas do parâmetro d para séries geradas de um processo ARFIMA(0, d , 0) não estacionário e níveis de contaminação 0% e 5%.

significativas para defasagens afastadas, além disso, o periodograma apresenta os maiores valores para frequências próximas de zero, sugerindo a presença da propriedade de memória longa nas observações. Para analisar o comportamento dos estimadores robustos em dados reais considera-se a série IGP-DI modificada, onde os valores correspondentes a fevereiro de 1999, outubro de 2002 e novembro de 2002 são substituídos pela média da série temporal (0.72%). A Figura 3 apresenta a série IGP-DI modificada e suas respectivas FAC, FACP e função periodograma. Para propósitos ilustrativos, são calculadas as estimativas do parâmetro de diferenciação fracionária, através dos estimadores definidos nas Eqs. 5, 9 e 13, para as séries original e modificada. Para o estimador definido na Eq. 13 são consideradas as janelas de Parzen (P), Bartlett (B), Tuckey-Hamming (TH) e Truncada (Tr). Com o intuito de analisar o comportamento dos estimadores robustos, a análise da série será limitada somente no cálculo das estimativas do parâmetro de diferenciação fracionária. A Tabela 8 mostra as estimativas obtidas a partir dos estimadores semi-paramétricos, com diferentes valores α definidos entre 0.5 e 0.8 e valor $\beta = 0.7$ para as janelas do estimador tipo L na Eq. 13.

Os resultados apresentados na Tabela 8 revelam que o estimador LP é afetado significativamente na presença dos dados substituídos pela média da série, tendendo a subestimar o verdadeiro valor do parâmetro d para qualquer valor do *bandwidth*. Esses resultados indicam que os valores identificados em fevereiro de 1999, outubro de 2002 e novembro de 2002 são possíveis valores atípicos. Para um α fixo, os estimadores robustos não apresentam variações significativas quando consideradas as séries original e modificada, sugerindo que os dados provêm de um processo estacionário com propriedade de memória longa com parâmetro próximo de $d = 0.3$. As diferenças nas estimativas obtidas com os estimadores tipo M e L podem ser explicadas pelo tamanho de amostra no conjunto de observações. Para a aplicação estimadores tipo L em séries reais, o tamanho de amostra deve ser maior, como sugerido nos experimentos de Monte Carlo na Seção 4.

5. COMENTÁRIOS FINAIS

Neste trabalho, propõe-se um estimador robusto do parâmetro de memória de processos ARFIMA baseado em uma variante do estimador LP proposto por Geweke & Porter-Hudak (1983), substituindo o periodograma clássico por um estimador robusto do espectro. Os estimadores robustos para o espectro são classificados em dois grupos: os estimadores tipo M e os estimadores tipo L .

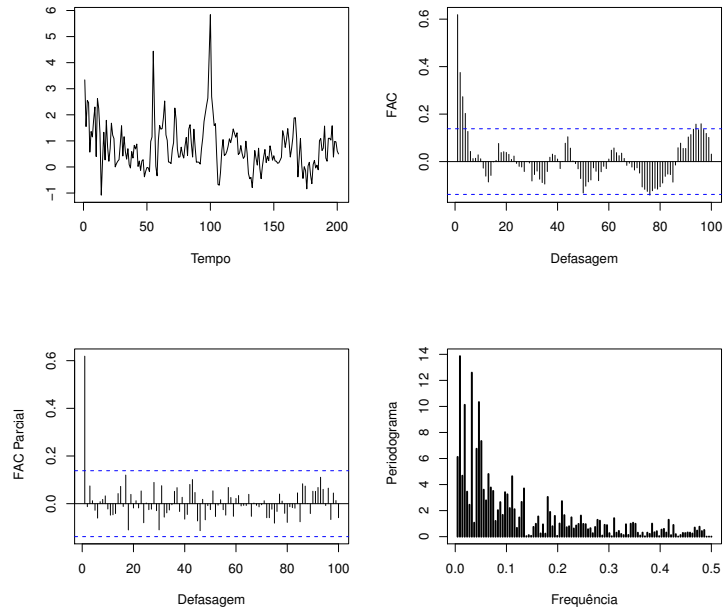


FIGURA 2. Série IGP-DI e Funções de autocorrelação: período: Ago/94 - Abr/11.

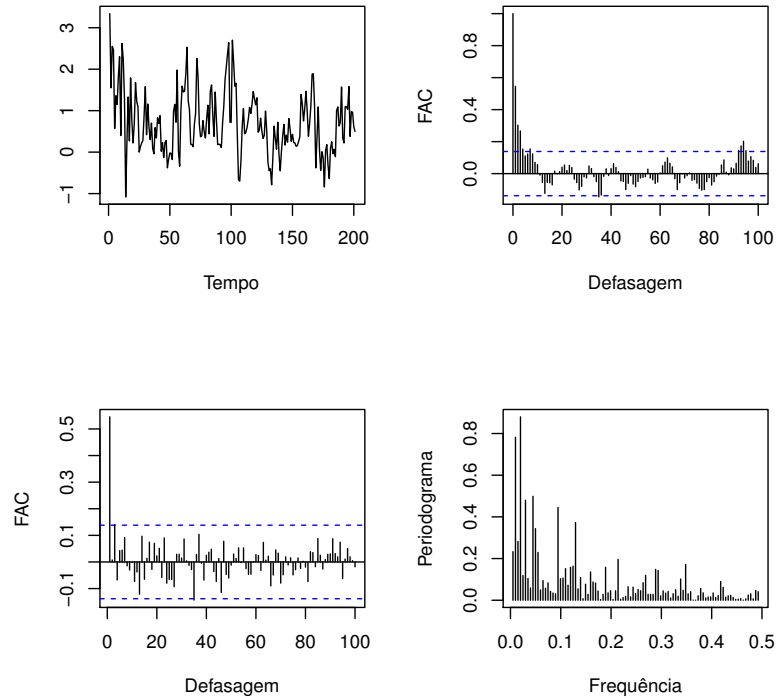


FIGURA 3. Série IGP-DI modificada e Funções de autocorrelação: período: Ago/94 - Abr/11.

A classe de estimadores tipo M baseia-se no cálculo de estimadores robustos dos parâmetros na representação espectral do processo sob análise. A estimativa dos parâmetros é realizada através de M -estimadores com diferentes funções de perda. Neste trabalho consideram-se as funções bisquare

Est	Série original				Série modificada			
	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.7$	$\alpha = 0.8$	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.7$	$\alpha = 0.8$
<i>LS</i>	0.0757 (0.3283)	0.1205 (0.1827)	0.3432 (0.1371)	0.3759 (0.0881)	0.3110 (0.1523)	0.3116 (0.1053)	0.3713 (0.0897)	0.3875 (0.0678)
<i>BS</i>	0.4415 (0.2371)	0.3853 (0.1497)	0.3150 (0.1022)	0.3163 (0.0796)	0.3492 (0.1996)	0.3418 (0.1242)	0.3248 (0.0863)	0.3678 (0.0889)
<i>HP</i>	0.2868 (0.1644)	0.3013 (0.1053)	0.3486 (0.0809)	0.4300 (0.0846)	0.3416 (0.1649)	0.3423 (0.1131)	0.3459 (0.0871)	0.3783 (0.0729)
<i>HU</i>	0.2663 (0.1747)	0.3216 (0.1110)	0.3484 (0.0884)	0.3890 (0.0826)	0.3187 (0.1851)	0.3461 (0.1200)	0.3199 (0.0853)	0.3428 (0.0709)
<i>P</i>	0.1802 (0.0857)	0.2335 (0.0745)	0.2269 (0.0469)	0.2397 (0.0331)	0.1630 (0.0782)	0.2077 (0.0603)	0.2078 (0.0385)	0.2230 (0.0251)
<i>TH</i>	0.1718 (0.0742)	0.1919 (0.0508)	0.2125 (0.0303)	0.2379 (0.0210)	0.1545 (0.0673)	0.1782 (0.0436)	0.1968 (0.0259)	0.2231 (0.0170)
<i>B</i>	0.1522 (0.0641)	0.1788 (0.0433)	0.2047 (0.0262)	0.2327 (0.0183)	0.1379 (0.0586)	0.1667 (0.0378)	0.1896 (0.0227)	0.2181 (0.0151)
<i>Tr</i>	0.1662 (0.0862)	0.2628 (0.0995)	0.2454 (0.0671)	0.2285 (0.0436)	0.1500 (0.0794)	0.2211 (0.0717)	0.2215 (0.0511)	0.2228 (0.0328)

TABELA 8. Estimativas d : Dados IGP-DI, período: Ago/94 - Abr/11.

(BS), Hampel (HP) e Huber (HU). A classe de estimadores tipo L baseia-se na utilização de um estimador robusto para a autocovariância no estimador alisado para o espectro definido na Eq. 13. O estimador robusto alisado é obtido com as *janelas* de Parzen, Bartlett, Tuckey-Hamming e truncada.

Resultados empíricos através de simulações de Monte Carlo sugerem um melhor desempenho para a classe de estimadores tipo M , quando utilizadas as funções *bisquare* e Hampel, em séries contaminadas por dados atípicos de tipo aditivo. Para os casos estacionário e não-estacionário, a classe de estimadores tipo M apresenta propriedades estatísticas equivalentes às do estimador LP, para tamanhos de amostra suficientemente grandes. Os resultados empíricos sugerem que essa classe de estimadores resulta uma alternativa razoável para estimação do parâmetro d em séries contaminadas.

Os estimadores tipo L apresentam menor desempenho, mesmo para tamanhos de amostra grande. A forma analítica dos estimadores tipo L provoca um aumento no custo computacional e uma diminuição significativa no desempenho estimador, favorecendo a utilização de estimadores tipo M para estimação do parâmetro de memória em processos fracionários. A classe de estimadores tipo L apresenta limitações para o cálculo de estimativas do parâmetro d na região de não-estacionariedade, o que sugere que o uso do mesmo deve ser evitado em situações onde não é possível garantir a estacionariedade da série.

Em síntese, o uso dos estimadores propostos apresenta uma vantagem significativa para a estimação do parâmetro de memória em processos ARFIMA, evitando o uso de metodologias alternativas para detecção de dados atípicos no conjunto de observações. Para situações práticas o uso dos estimadores tipo M resulta uma opção razoável para estimação do parâmetro de diferenciação fracionária de processos estacionários e não-estacionários. A análise empírica mostra que, entre todas as funções de perda aqui consideradas, as funções *bisquare* e Hampel são as melhores opções para obtenção de estimativas robustas do d em séries na presença de contaminação.

O estudo de metodologias robustas na análise de séries temporais é uma vertente de pesquisa pouco explorada na atualidade. O uso de tais metodologias na modelagem de séries temporais evita procedimentos para detecção de *outliers* nas observações, sendo esta uma vertente de pesquisa de grande interesse para pesquisadores da área. Trabalhos futuros incluem o desenvolvimento de metodologias robustas, baseadas nos periodogramas robustos, para estimação dos coeficientes autorregressivos e de médias móveis em processos ARFIMA contaminados, assim como a estimação dos parâmetros de diferenciação fracionária sazonais e o desenvolvimento de estimadores robustos para o parâmetro de memória em processos ARFIMA multivariados.

PROVAS

Prova do Teorema 1. A transformada discreta de Fourier do processo contaminado é dada por

$$\begin{aligned} w_z(\lambda) &= \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n z_t e^{i\lambda t} = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n (y_t + v_t) e^{i\lambda t} \\ &= \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n y_t e^{i\lambda t} + \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n v_t e^{i\lambda t} = w_y(\lambda) + w_v(\lambda). \end{aligned} \quad (15)$$

Avaliando nas frequências de Fourier e multiplicando pelo conjugado tem-se que

$$\begin{aligned} I_{zj} &:= \{w_y(\lambda_j) + w_v(\lambda_j)\} \overline{\{w_y(\lambda_j) + w_v(\lambda_j)\}} \\ &= I_{yj} + I_{vj} + 2\Re\{w_{yj}\overline{w_{vj}}\}, \end{aligned}$$

onde a TDF do processo $\{y_t; t = 1, 2, \dots, n\}$ é dada por

$$\begin{aligned} w_{yj} &= \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n y_t e^{i\lambda_j t} = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n \left(c + \sum_{k=1}^t x_k \right) e^{i\lambda_j t} \\ &= \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n \sum_{k=1}^t x_k e^{i\lambda_j t}, \end{aligned} \quad (16)$$

onde $D(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n e^{i\lambda_j t} = 0$. A TDF do processo $\{x_t; t = 1, 2, \dots, n\}$ é dada por

$$\begin{aligned} w_{xj} &= \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n x_t e^{i\lambda_j t} = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n (y_t - y_{t-1}) e^{i\lambda_j t} \\ &= \frac{1}{\sqrt{2\pi n}} \left[\sum_{t=1}^n y_t e^{i\lambda_j t} - e^{i\lambda_j} \sum_{t=1}^n y_t e^{i\lambda_j t} + (y_n e^{i\lambda_j(n+1)} - y_0 e^{i\lambda_j}) \right] \\ &= \frac{1}{\sqrt{2\pi n}} \left[(1 - e^{i\lambda_j}) \sum_{t=1}^n y_t e^{i\lambda_j t} + (y_n e^{i\lambda_j} - y_0 e^{i\lambda_j}) \right] \\ &= (1 - e^{i\lambda_j}) w_{yj} + \frac{e^{i\lambda_j}}{\sqrt{2\pi n}} y_n. \end{aligned}$$

Multiplicando pelo conjugado obtém-se o resultado. \square

REFERÊNCIAS

- Agostinelli, C. & Bisaglia, L. (2003), Robust estimation of arfima processes, Technical report, Università Ca' Foscari di Venezia.
- Arcones, M. (1998), 'Asymptotic theory for M-estimates over a convex kernel', *Econometric Theory* **14**, 387–422.
- Bai, Z. D., Rao, C. R. & Wu, Y. (1992), 'M-estimates of multivariate linear regression parameters under a convex discrepancy function', *Statistica sinica* **2**, 237–254.
- Bai, Z. D., Rao, C. R. & Wu, Y. H. (1997), M-estimation of multivariate linear regression by minimizing the difference of two convex functions, in 'Handbook of Statistics', Vol. 15, Elsevier Science, North-Holland, pp. 1–19.
- Beltrão, K. & Bloomfield, P. (1987), 'Determining the bandwidth of a kernel spectrum estimate', *Journal of Time Series Analysis* **8**, 21–38.
- Beran, J. (1994), 'On a class of M-estimators for gaussian long-memory models', *Biometrika* **81**, 755–766.

- Chan, W.-S. (1992), ‘A note on time series model specification in the presence outliers’, *Journal of Applied Statistics* **19**, 117–124.
- Chan, W.-s. (1995), ‘Outliers and financial time series modelling: a cautionary note’, *Mathematics and Computers in Simulation* **39**, 425–430.
- Chang, I., Tiao, G. C. & Chen, C. (1988), ‘Estimation of time series parameters in presence of outliers’, *Technometrics* **30**, 1936–204.
- Chen, C. & Liu, L.-M. (1993), ‘Joint estimation of model parameters and outlier effects in time series’, *Journal of the American Statistical Association* **88**, 284–297.
- Chen, G., Abraham, B. & Peiris, S. (1994), ‘Lag window estimation of the degree of differencing in fractionally integrated time series models’, *Journal of Time Series Analysis* **15**, 473–487.
- Deutsch, S. J., Richards, J. E. & Swain, J. J. (1990), ‘Effects of a single outlier on arma identification’, *Communications in Statistics: Theory and Methods* **19**, 2207–2227.
- Fajardo, F. (2011), ‘On some alternatives for robust estimation of the spectrum in stationary processes’, *Brazilian Review of Econometrics* (available in <http://bibliotecadigital.fgv.br/ojs/index.php/bre/article/view/2767>) .
- Fajardo, F. A. & Reisen, V. A. (2011), ‘Estimadores robustos para o espectro de processos estacionários’, *In press* .
- Fajardo, F., Reisen, V. A. & Cribari-Neto, F. (2009), ‘Robust estimation in long-memory processes under additive outliers’, *Journal of Statistical Planning and Inference* **139**, 2511–2525.
- Fox, R. & Taqqu, M. S. (1986), ‘Large-sample properties of parameters estimates for strongly dependent stationary gaussian time series’, *The Annals of Statistics* **14**, 517–532.
- Geweke, J. & Porter-Hudak, S. (1983), ‘The estimation and application of long memory time series model’, *Journal of Time Series Analysis* **4**, 221–238.
- Granger, C. W. J. & Joyeux, R. (1980), ‘An introduction to long-memory time series models and fractional differencing’, *Journal of Time Series Analysis* **1**, 15–30.
- Haldrup, N. & Nielsen, M. O. (2007), ‘Estimation of fractional integration in the presence of data noise’, *Computational Statistics & Data Analysis* **51**, 3100–3114.
- Hampel, F., Ronchetti, E., Rousseeuw, P. & Stahel, W. (1986), *Robust statistics: The approach based on influence functions*, John Wiley & sons.
- Hannan, E. J. (1973), ‘Central limit theorems for time series regression’, *Probability theory and related fields* **26**, 157–170.
- Hosking, J. R. (1981), ‘Fractional differencing’, *Biometrika* **68**, 165–176.
- Huber, P. (1964), ‘Robust estimation of a location parameter’, *The Annals of Mathematical Statistics* **35**, 73–101.
- Huber, P. J. (2004), *Robust Statistics*, third edn, John Wiley & Sons.
- Hurvich, C. & Beltrão, K. (1990), ‘Cross-validatory choice of a spectrum estimate and its connections with aic’, *Journal of Time Series Analysis* **11**, 121–137.
- Hurvich, C. M. & Beltrão, K. I. (1993), ‘Asymptotics for low-frequency ordinates of the periodogram of a long-memory time series’, *Journal of Time Series Analysis* **14**(5), 455–472.
- Hurvich, C. M., Deo, R. & Brodsky, J. (1998), ‘The mean square error of geweke and porter-hudaks estimator of the memory parameter of a long-memory time series’, *Journal of Time Series Analysis* **19**(1), 19–46.
- Hurvich, C. M. & Ray, B. K. (1995), ‘Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes’, *Journal of Time Series Analysis* **16**(1), 17–42.
- Kim, C. S. & Phillips, P. (2006), Log periodogram regression: the nonstationary case, Technical report, Cowles Foundation Discussion Paper, Yale University.
- Künsch, H. R. (1987), Statistical aspects of self-similar processes, in Y. Prohorov & V. Sazanov, eds, ‘Proceedings of the first world congress of the Bernoulli society’, Vol. 1, Utrecht, pp. 67–74.

- Lopes, S. R. C., Olbermann, B. P. & Reisen, V. A. (2004), ‘Comparison of estimation methods in non-stationary arfima process’, *Journal of Statistical Computation and Simulation* **74**(5), 339–347.
- Ma, Y. & Genton, M. (2000), ‘Highly robust estimation of the autocovariance function’, *Journal of Time Series Analysis* **21**, 663–684.
- Maronna, R., Martin, R. D. & Yohai, V. (2006), *Robust statistics*, John Wiley & sons.
- Phillips, P. (1999), Discrete fourier transforms of fractional processes, Technical report, Cowles Foundation Discussion Paper, Yale University.
- Phillips, P. (2007), ‘Unit root log periodogram regression’, *Journal of econometrics* **138**, 104–124.
- Reisen, V. A. (1994), ‘Estimation of the fractional difference parameter in the arima(p, d, q) model using the smoothed periodogram’, *Journal of Time Series Analysis* **15**, 335–350.
- Robinson, P. (1994), ‘Efficient tests of nonstationary hypotheses’, *Journal of the American Statistical Association* **89**, 1420–1437.
- Robinson, P. (1996), ‘Optimal spectral bandwidth for long memory’, *Statistica sinica* **6**, 97–112.
- Robinson, P. M. (1995a), ‘Gaussian semiparametric estimation of long range dependence’, *The annals of statistics* **23**, 1630–1661.
- Robinson, P. M. (1995b), ‘Log-periodogram regression of time series with long range dependence’, *The annals of statistics* **23**, 1048–1072.
- Rousseeuw, P. J. & Croux, C. (1993), ‘Alternatives to the median absolute deviation’, *Journal of the American Statistical Association* **88**, 1273–1283.
- Shimotsu, K. & Phillips, P. (2005), ‘Exact local whittle estimation of fractional integration’, *Annals of statistics* **33**, 1890–1933.
- Tanaka, K. (1999), ‘The nonstationary fractional unit root’, *Econometric Theory* **15**, 549–582.
- Velasco, C. (1999a), ‘Gaussian semiparametric estimation of non-stationary time series’, *Journal of Time Series Analysis* **20**(1), 87–127.
- Velasco, C. (1999b), ‘Non-stationary log-periodogram regression’, *Journal of Econometrics* **91**, 325–371.
- Wu, W. B. (2007), ‘M-estimation of linear models with dependent errors’, *The annals of statistics* **35**, 495–521.
- Yajima, Y. (1989), ‘A central limit theorem for fourier transform of strongly dependent stationary’, *journal of time series analysis* **10**, 375–383.