Physica A 406 (2014) 169–175

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Detrending moving-average cross-correlation coefficient: Measuring cross-correlations between non-stationary series

Ladislav Kristoufek *

Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodarenskou Vezi 4, 182 08, Prague 8, Czech Republic
Institute of Economic Studies, Faculty of Social Sciences, Charles University, Opletalova 26, 110 00, Prague 1, Czech Republic

HIGHLIGHTS

- DMCA coefficient is introduced.
- Different settings (non-stationarity level, scales, correlations, time series length) are examined.
- DMCA coefficient is both an alternative and a complement to the DCCA coefficient.

ARTICLE INFO

Article history:
Received 4 November 2013
Received in revised form 14 January 2014
Available online 26 March 2014

Keywords:
Correlations
Econophysics
Non-stationarity

ABSTRACT

In the paper, we introduce a new measure of correlation between possibly non-stationary series. As the measure is based on the detrending moving-average cross-correlation analysis (DMCA), we label it as the DMCA coefficient $\rho_{DMCA}(\lambda)$ with a moving average window length λ. We analytically show that the coefficient ranges between -1 and 1 as a standard correlation does. In the simulation study, we show that the values of $\rho_{DMCA}(\lambda)$ very well correspond to the true correlation between the analyzed series regardless the (non-)stationarity level. Dependence of the newly proposed measure on other parameters – correlation level, moving average window length and time series length – is discussed as well.

1. Introduction

Inspection of statistical properties of multivariate series has become a topic of increasing importance in econophysics. For this purpose, various estimators of power-laws in cross-correlations of a pair of series have been proposed—detrended cross-correlation analysis and its various versions [1,2], detrending cross-correlation moving average [3,4], height cross-correlation analysis [5] and cross-correlation analysis based on statistical moments [6]. Out of these, the detrended cross-correlation analysis (DCCA) [1,2] has become the most popular one. Apart from the analysis of the power laws in the cross-correlation function itself, Zebende [7] proposed the DCCA cross-correlation coefficient as a combination of DCCA and the detrended fluctuation analysis (DFA) [8–10]. Even though its ability to uncover power-law cross-correlations has been somewhat disputed [11–14], Kristoufek [15] shows that the coefficient is able to estimate the correlation coefficient between non-stationary series precisely and that it dominates the standardly used Pearson's correlation coefficient.

* Corresponding author at: Institute of Economic Studies, Faculty of Social Sciences, Charles University, Opletalova 26, 110 00, Prague 1, Czech Republic. Tel.: +420 607684417.
E-mail addresses: kristoufek@ies-prague.org, kristouf@utia.cas.cz.

http://dx.doi.org/10.1016/j.physa.2014.03.015
0378-4371/

Historically, the detrended fluctuation analysis and methods derived from it have been frequently compared to (or sometimes even interwoven with) the detrending moving average (DMA) [16,17] procedures. In most cases, the competing approaches fare very similarly while the DMA algorithms are computationally less demanding as they contain no box-splitting[1] and regression fitting [18–26]. In this paper, we follow these steps and propose an alternative but also a complementary coefficient to the DCCA cross-correlation coefficient of Zebende [7]—the detrending moving-average cross-correlation coefficient. In the following section, the coefficient is introduced. After that, results of the wide Monte Carlo study are presented showing that the newly proposed coefficient estimates the true correlation coefficient precisely even for strongly non-stationary series. Comparison to the DCCA coefficient is included as well.

2. DMCA coefficient

We start with the detrending moving average procedure (DMA) proposed by Vandewalle & Ausloos [16] and further developed by Alessio et al. [17]. For (possibly asymptotically non-stationary) series $\{x_t\}$ and $\{y_t\}$, we construct integrated series $X_t = \sum_{i=1}^{t} x_i$ and $Y_t = \sum_{i=1}^{t} y_i$ for $t = 1, 2, \ldots, T$ where T is the time series length which is common for both series. Fluctuation functions $F_{x,DMA}$ and $F_{y,DMA}$ are then defined as

$$F^2_{x,DMA}(\lambda) = \frac{1}{T-\lambda+1} \sum_{i=\lfloor \lambda-\theta(\lambda-1)\rfloor}^{\lfloor T-\theta(\lambda-1)\rfloor} \left(X_t - \widetilde{X_{t,\lambda}}\right)^2, \tag{1}$$

$$F^2_{y,DMA}(\lambda) = \frac{1}{T-\lambda+1} \sum_{i=\lfloor \lambda-\theta(\lambda-1)\rfloor}^{\lfloor T-\theta(\lambda-1)\rfloor} \left(Y_t - \widetilde{Y_{t,\lambda}}\right)^2 \tag{2}$$

where λ is the moving average window length and θ is a factor of moving average type (forward, centered and backward for $\theta = 0, \theta = 0.5$ and $\theta = 1$, respectively). $\widetilde{X_{t,\lambda}}$ and $\widetilde{Y_{t,\lambda}}$ then represent the specific moving averages with the window size λ at time t. Different types of moving averages have been studied and the centered one ($\theta = 0.5$) shows the best results [18] so that we apply $\theta = 0.5$ in this study as well.

For the bivariate series, He & Chen [4] propose the detrending moving-average cross-correlation analysis (DMCA) which is a special case of the method proposed by Arianos & Carbone [3]. The bivariate fluctuation F^2_{DMCA}, which can be seen as a detrended covariance, is defined as

$$F^2_{DMCA}(\lambda) = \frac{1}{T-\lambda+1} \sum_{i=\lfloor \lambda-\theta(\lambda-1)\rfloor}^{\lfloor T-\theta(\lambda-1)\rfloor} \left(X_t - \widetilde{X_{t,\lambda}}\right)\left(Y_t - \widetilde{Y_{t,\lambda}}\right). \tag{3}$$

In the steps of Zebende [7], we propose the detrending moving-average cross-correlation coefficient, or also the DMCA-based correlation coefficient, as

$$\rho_{DMCA}(\lambda) = \frac{F^2_{DMCA}(\lambda)}{F_{x,DMA}(\lambda)F_{y,DMA}(\lambda)}. \tag{4}$$

In a similar way as for the DCCA correlation coefficient [11], the DMCA coefficient can be rewritten as

$$\begin{aligned}\rho_{DMCA}(\lambda) &= \frac{F^2_{DMCA}(\lambda)}{F_{x,DMA}(\lambda)F_{y,DMA}(\lambda)} \\ &= \frac{\frac{1}{T-\lambda+1}\sum_{i=\lfloor \lambda-\theta(\lambda-1)\rfloor}^{\lfloor T-\theta(\lambda-1)\rfloor}\left(X_t - \widetilde{X_{t,\lambda}}\right)\left(Y_t - \widetilde{Y_{t,\lambda}}\right)}{\sqrt{\frac{1}{T-\lambda+1}\sum_{i=\lfloor \lambda-\theta(\lambda-1)\rfloor}^{\lfloor T-\theta(\lambda-1)\rfloor}\left(X_t - \widetilde{X_{t,\lambda}}\right)^2 \frac{1}{T-\lambda+1}\sum_{i=\lfloor \lambda-\theta(\lambda-1)\rfloor}^{\lfloor T-\theta(\lambda-1)\rfloor}\left(Y_t - \widetilde{Y_{t,\lambda}}\right)^2}} \\ &= \frac{\sum_{i=\lfloor \lambda-\theta(\lambda-1)\rfloor}^{\lfloor T-\theta(\lambda-1)\rfloor}\epsilon_{x,t}\epsilon_{y,t}}{\sqrt{\sum_{i=\lfloor \lambda-\theta(\lambda-1)\rfloor}^{\lfloor T-\theta(\lambda-1)\rfloor}\epsilon^2_{x,t}\sum_{i=\lfloor \lambda-\theta(\lambda-1)\rfloor}^{\lfloor T-\theta(\lambda-1)\rfloor}\epsilon^2_{y,t}}}\end{aligned} \tag{5}$$

where $\{\epsilon_{x,t}\}$ and $\{\epsilon_{y,t}\}$ are the series $\{X_t\}$ and $\{Y_t\}$, respectively, detrended by the centered moving average of length λ. From the last part of Eq. (5), it is visible that

$$-1 \leq \rho_{DMCA}(\lambda) \leq 1 \tag{6}$$

[1] Note that moving averages are also utilized in the detrended fluctuation analysis methods where the polynomial detrending is substituted by the moving average filtering [25]. However, DMA presented in this paper is based on scaling of fluctuations with moving average window length whereas DFA methods using moving averages are still based on box splitting and scaling with box sizes.

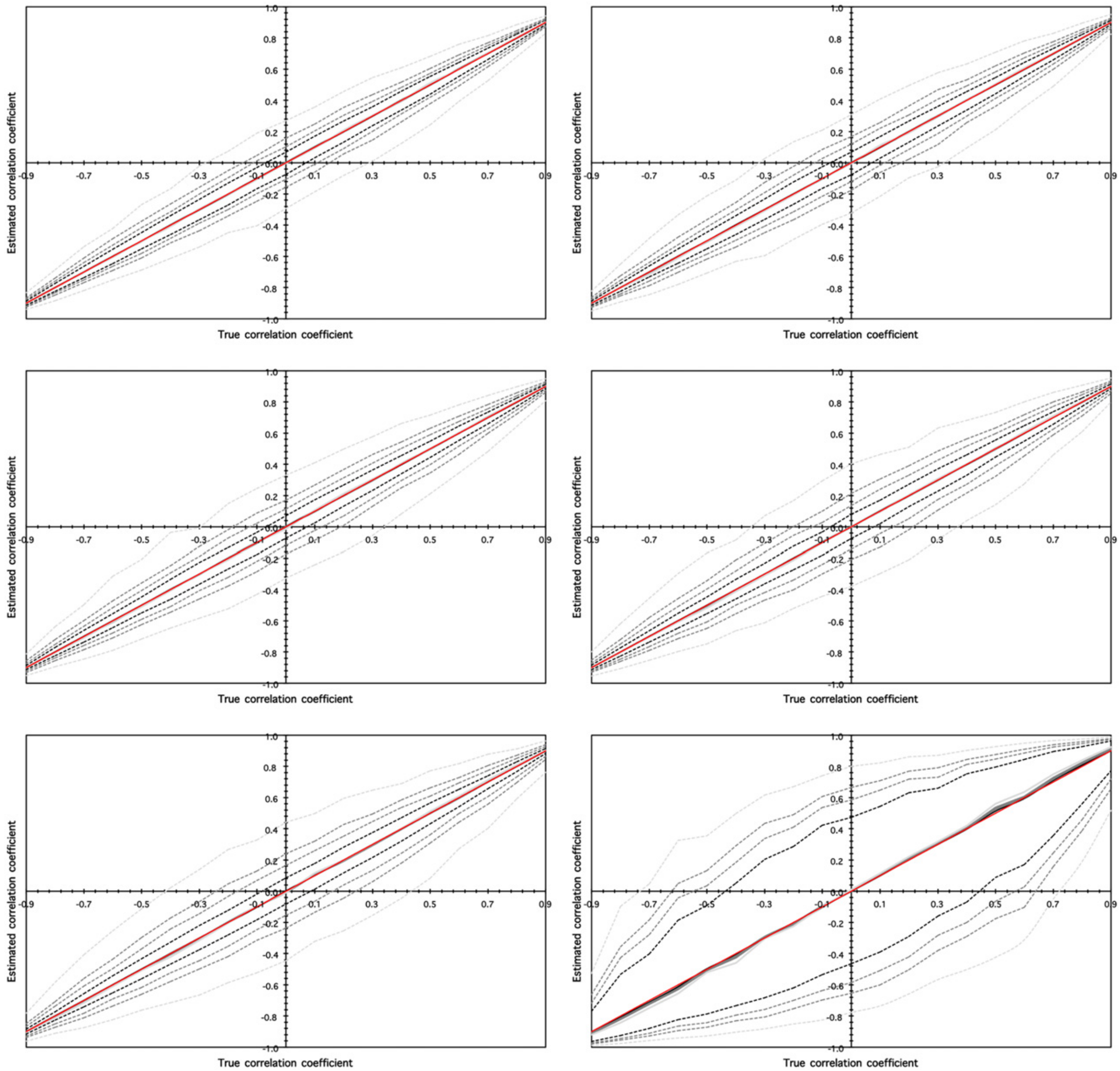

Fig. 1. *Estimated DMCA correlation coefficients for different fractional integration parameters d I.* Results for the time series of length $T = 1000$ are shown here. Separate figures represent different parameters $d - d = 0.1$ (top left), $d = 0.4$ (top right), $d = 0.6$ (middle left), $d = 0.9$ (middle right), $d = 1.1$ (bottom left), $d = 1.4$ (bottom right). Red lines represent the true value of $\rho_{\varepsilon\nu}$. The solid lines of shades of gray (mostly overlapping with the red line) represent the median values of 1000 simulations for the given parameter setting. The dashed lines represent the 95% confidence intervals (the 2.5th and the 97.5th quantiles of the simulations). Different shades of gray stand for different values of the moving average window λ going from the lowest one ($\lambda = 5$, the darkest shade) to the highest one ($\lambda = 101$, the lightest shade). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

according to the Cauchy–Schwarz inequality. The DMCA-based correlation coefficient thus has the same range as the standard correlation. In the following section, we show that the values of the newly proposed coefficient very precisely describe the correlations between two series and this is true even for strongly non-stationary series.

3. Simulations results

In this section, we show that the DMCA coefficient is able to describe correlations between (even non-stationary) series very precisely. To do so, we present a wide Monte Carlo simulation study[2] for varying level of correlations and

[2] The R-project codes for the DMCA coefficient are available at http://staff.utia.cas.cz/kristoufek/Ladislav_Kristoufek/Codes.html or upon request from the author.

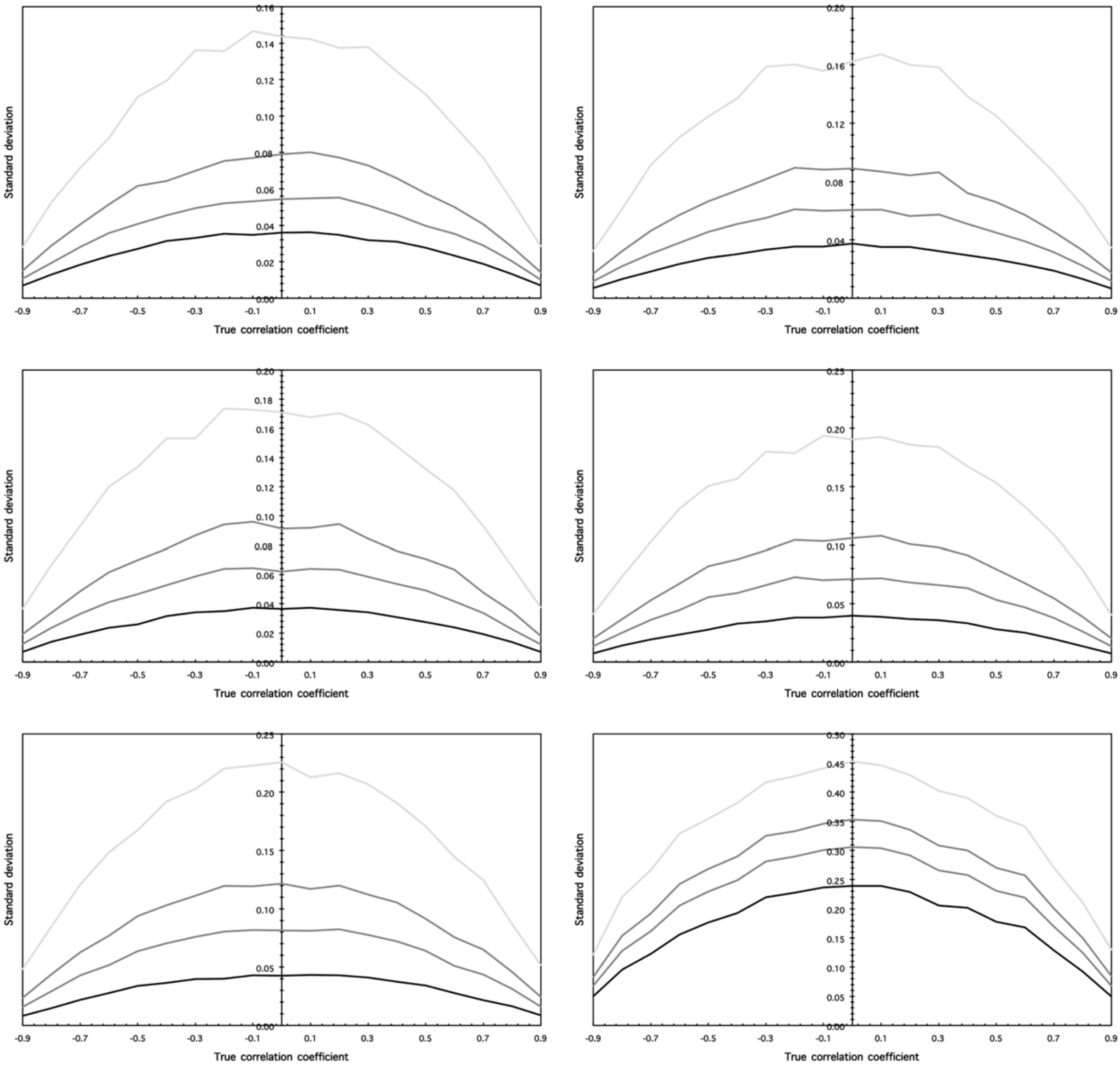

Fig. 2. *Standard deviations of DMCA correlation coefficients for different fractional integration parameters d I.* Results for the time series of length $T = 1000$ are shown here. Separate figures represent different parameters $d - d = 0.1$ (top left), $d = 0.4$ (top right), $d = 0.6$ (middle left), $d = 0.9$ (middle right), $d = 1.1$ (bottom left), $d = 1.4$ (bottom right). Solid lines represent the standard deviation of 1000 simulations for given parameter setting. Different shades of gray stand for different values of the moving average window λ going from the lowest one ($\lambda = 5$, the darkest shade) to the highest one ($\lambda = 101$, the lightest shade).

(non-)stationarity. For this purpose, we utilize two ARFIMA(0, d, 0) processes with correlated error terms

$$x_t = \sum_{n=0}^{\infty} a_n(d_1)\varepsilon_{t-n} \tag{7}$$

$$y_t = \sum_{n=0}^{\infty} a_n(d_2)\nu_{t-n} \tag{8}$$

where $a_n(d_i) = \frac{\Gamma(n+d_i)}{\Gamma(n+1)\Gamma(d_i)}$, $\langle\varepsilon_t\rangle = \langle\nu_t\rangle = 0$, $\langle\varepsilon_t^2\rangle = \langle\nu_t^2\rangle = 1$ and $\langle\varepsilon_t\nu_t\rangle = \rho_{\varepsilon\nu}$.

Parameter d is important for the stationarity discussion. For $d < 0.5$, the series are stationary while for $0.5 \leq d < 1$, the series are non-stationary but mean-reverting whereas for $d \geq 1$, the series are non-stationary and non-mean-reverting (explosive). To see how the DMCA coefficient is able to quantify the correlation for (non-)stationary series, we select several

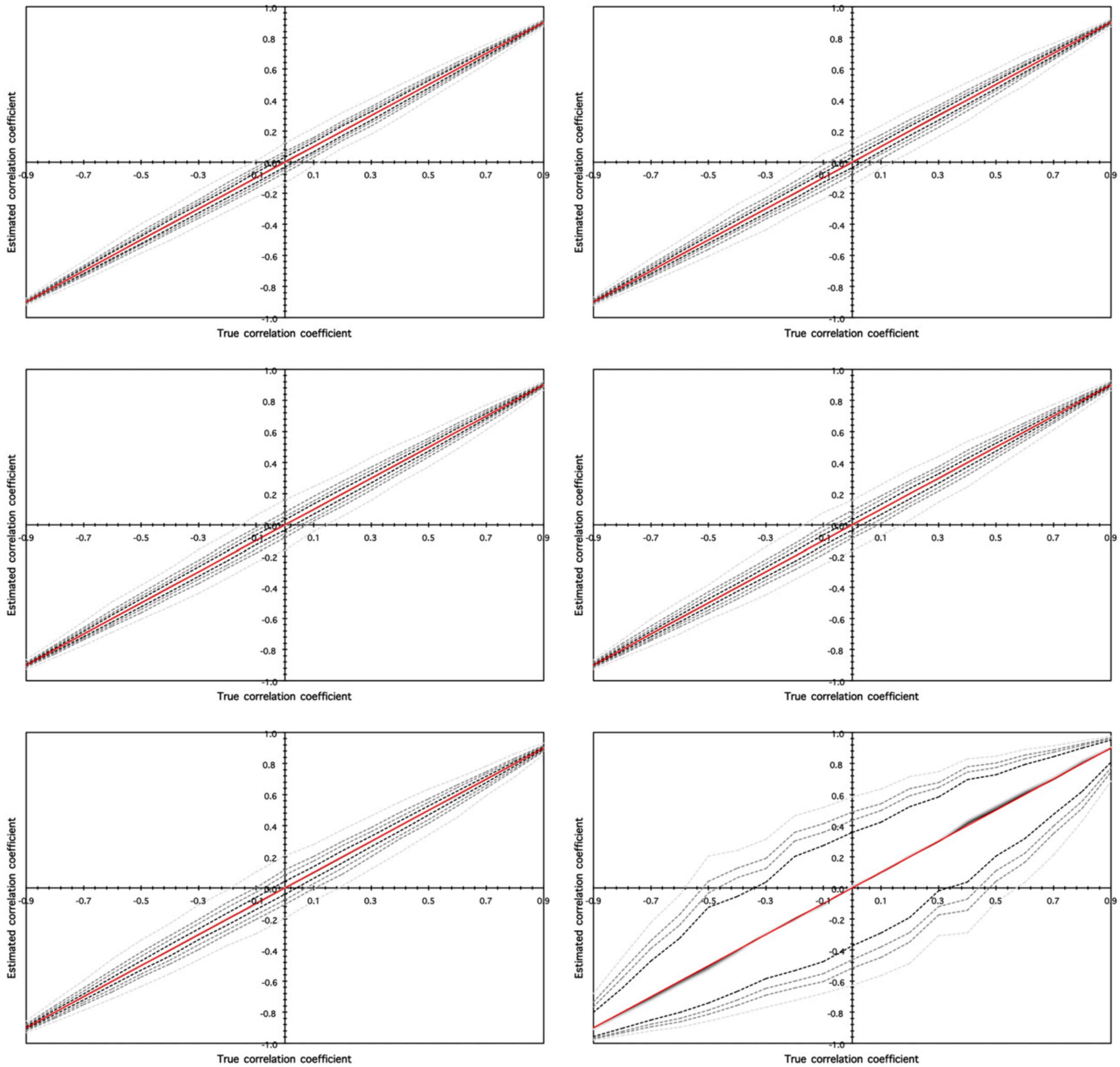

Fig. 3. *Estimated DMCA correlation coefficients for different fractional integration parameters d II.* Results for the time series of length $T = 5000$ are shown here. Notation of Fig. 1 is used.

levels of d, specifically $d_1 = d_2 \equiv d = 0.1, 0.4, 0.6, 0.9, 1.1, 1.4$. To cover a wide spectrum of possible correlation levels in a sufficient detail, we study $\rho_{\varepsilon\nu} = -0.9, -0.8, \ldots, 0.8, 0.9$. To see the effect of different moving average lengths, we study the cases $\lambda = 5, 15, 31, 101$. And finally, we examine two time series lengths which are representative for usually analyzed series in econophysics—$T = 1000, 5000$.

In Figs. 1–4, all the results are summarized. In Figs. 1–2, the time series length of $T = 1000$ is discussed, and in Figs. 3–4, the results for $T = 5000$ are illustrated. In the figures, we present the 2.5%, 50% and 97.5% quantiles, i.e. we show the 95% confidence intervals and the median value, based on 1000 simulations for a given parameter setting.

The main findings can be summarized as follows. First, the DMCA coefficient is an unbiased estimator of the true correlation coefficient of the series regardless the (non-)stationarity setting, the correlation level, the time series length and the moving average window size λ. Second, the confidence intervals get wider with an increasing λ. Nonetheless, the confidence intervals remain quite narrow for all inspected λs (when compared to the DCCA coefficient or Pearson's correlation examined by Kristoufek [15]). Third, the performance of the coefficient is symmetric around the zero correlation, i.e. there are no evident differences between the performance for the positive and for the negative correlations. Fourth, the DMCA coefficient gets more precise with an increasing absolute value of the true correlation—for the true correlation of both ± 0.9, the confidence intervals are extremely narrow. Fifth, the standard deviation of the coefficient is approximately

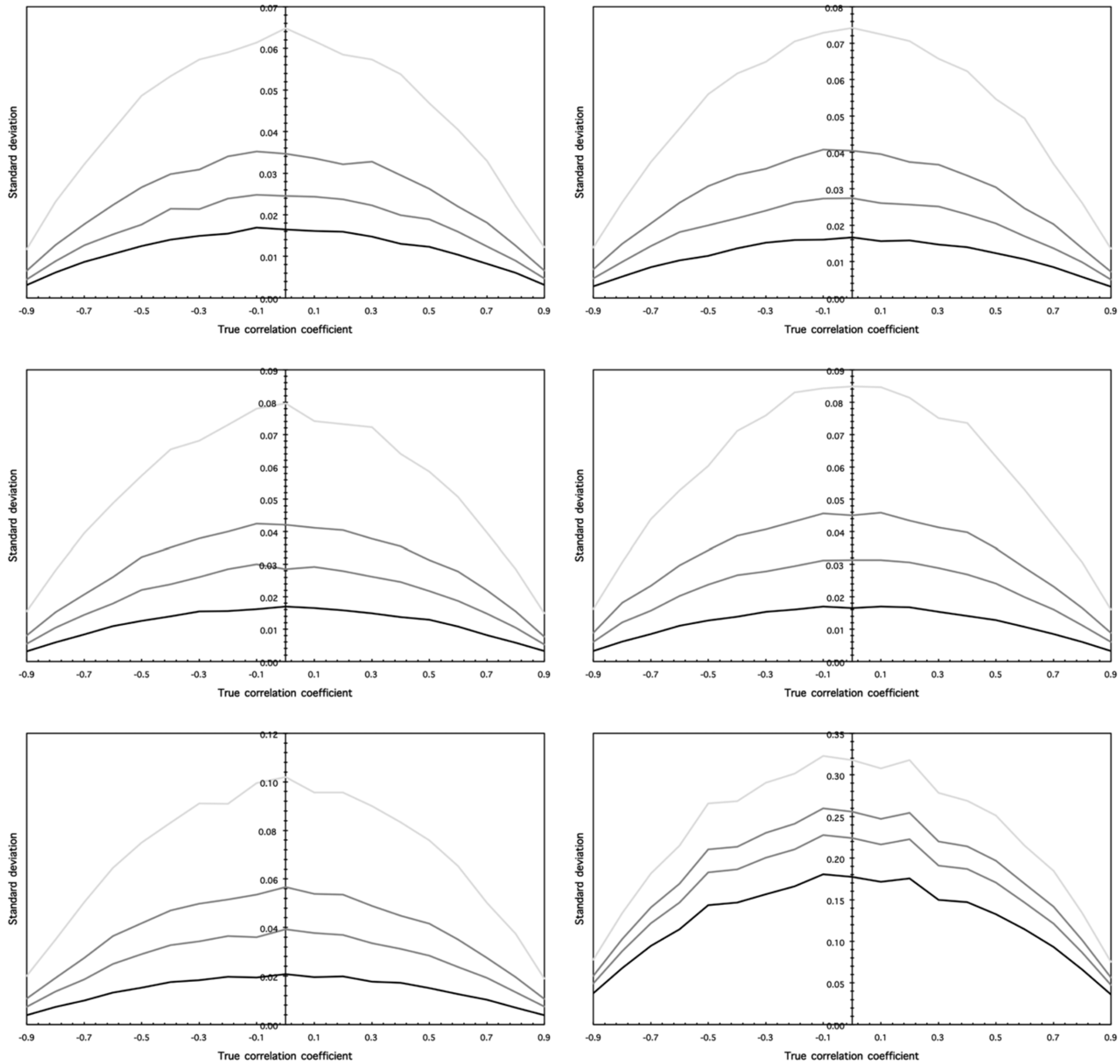

Fig. 4. *Standard deviations of DMCA correlation coefficients for different fractional integration parameters d II.* Results for the time series of length $T = 5000$ are shown here. Notation of Fig. 2 is used.

symmetric around the zero correlation and it increases with the parameter d. And sixth, the performance of the DMCA coefficient gets better with an increasing time series length T.

4. Conclusions and discussion

In this paper, we introduce a new measure of correlation between possibly non-stationary series. As the measure is based on the detrending moving-average cross-correlation analysis (DMCA), we label it as the DMCA coefficient $\rho_{DMCA}(\lambda)$ with a moving average window size λ. We analytically show that the coefficient ranges between -1 and 1 as a standard correlation does. In the simulation study, we show that the values of $\rho_{DMCA}(\lambda)$ very well correspond to the true correlation between the analyzed series regardless the (non-)stationarity level.

As the $\rho_{DMCA}(\lambda)$ coefficient can be seen as both an alternative and a complement to the $\rho_{DCCA}(s)$ coefficient [7], the precision of these coefficients should be compared. As both this study and our previous study of the statistical properties of the DCCA coefficient [15] are constructed in a similar manner, the comparison is easy. The most important findings are the following. First, both coefficients provide an unbiased estimator even for highly non-stationary processes. Second, apart from the very strong non-stationarity case ($d = 1.4$), both coefficients bring very precise estimates of the true correlation. And third, turning to the differences, we find that the DMCA coefficient has narrower confidence intervals and lower standard deviations for given settings. However, it needs to be noted that the crucial parameters of the coefficients are different. For the DMCA method, we have the moving average window length λ, and for the DCCA method, we utilize the scale s which is used for box-splitting and averaging of the fluctuations around the time trend. Therefore, these two coefficients are not easily comparable. However, an indisputable advantage of the newly proposed DMCA coefficient remains—the DMCA procedure is computationally much less demanding than the DCCA method mainly due to the box-splitting and the time-trend fitting in the DCCA procedure. This might become an issue for long financial series and in any other branch of research. Overall, we suggest to use both methods in empirical analyses dealing with potentially non-stationary series as each of the methods can better control for different types of trends.

Acknowledgments

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007–2013) under grant agreement No. SSH.2013.1.3-2-612955 (FinMaP). Support from the Czech Science Foundation under projects No. P402/11/0948 and No. 14-11402P is also gratefully acknowledged.

References

[1] B. Podobnik, H.E. Stanley, Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series, Phys. Rev. Lett. 100 (2008) 084102.
[2] W.-X. Zhou, Multifractal detrended cross-correlation analysis for two nonstationary signals, Phys. Rev. E 77 (2008) 066211.
[3] S. Arianos, A. Carbone, Cross-correlation of long-range correlated series, J. Stat. Mech. Theory Exp. 3 (2009) P03037.
[4] L.-Y. He, S.-P. Chen, A new approach to quantify power-law cross-correlation and its application to commodity markets, Physica A 390 (2011) 3806–3814.
[5] L. Kristoufek, Multifractal height cross-correlation analysis: A new method for analyzing long-range cross-correlations, EPL 95 (2011) art. 68001.
[6] J. Wang, P. Shang, W. Ge, Multifractal cross-correlation analysis based on statistical moments, Fractals 20 (3) (2012) 271–279.
[7] G.F. Zebende, DCCA cross-correlation coefficient: quantifying level of cross-correlation, Physica A 390 (2011) 614–618.
[8] C. Peng, S. Buldyrev, A. Goldberger, S. Havlin, M. Simons, H.E. Stanley, Finite-size effects on long-range correlations: Implications for analyzing DNA sequences, Phys. Rev. E 47 (5) (1993) 3730–3733.
[9] C. Peng, S. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A. Goldberger, Mosaic organization of DNA nucleotides, Phys. Rev. E 49 (2) (1994) 1685–1689.
[10] J. Kantelhardt, S. Zschiegner, E. Koscielny-Bunde, A. Bunde, S. Havlin, H.E. Stanley, Multifractal detrended fluctuation analysis of nonstationary time series, Physica A 316 (1–4) (2002) 87–114.
[11] B. Podobnik, Z.-Q. Jiang, W.-X. Zhou, H.E. Stanley, Statistical tests for power-law cross-correlated processes, Phys. Rev. E 84 (2011) 066118.
[12] R. Balocchi, M. Varanini, A. Macerate, Quantifying different degrees of coupling in detrended cross-correlation analysis, EPL 101 (2013) 20011.
[13] D. Blythe, A Rigorous and Efficient Asymptotic Test for Power-law Cross-correlation. Technical Report, 2013, arXiv:1309.4073.
[14] G.F. Zebende, M.F. da Silva, A. Machado Filho, DCCA cross-correlation coefficient differentiation: theoretical and practical approaches, Physica A 392 (2013) 1756–1761.
[15] L. Kristoufek, Measuring correlations between non-stationary series with DCCA coefficient, Physica A 402 (2014) 291–298.
[16] N. Vandewalle, M. Ausloos, Crossing of two mobile averages: a method for measuring the roughness exponent, Phys. Rev. E 58 (1998) 6832–6834.
[17] E. Alessio, A. Carbone, G. Castelli, V. Frappietro, Second-order moving average and scaling of stochastic time series, Eur. Phys. J. B 27 (2002) 197–200.
[18] A. Carbone, G. Castelli, Scaling properties of long-range correlated noisy signals: application to financial markets, Proc. SPIE 5114 (2003) 406–414.
[19] L. Xu, P. Ivanov, K. Hu, Z. Chen, A. Carbone, H.E. Stanley, Quantifying signals with power-law correlations: a comparative study of detrended fluctuation analysis and detrending moving average techniques, Phys. Rev. E 71 (2005) 051101.
[20] D. Grech, Z. Mazur, Statistical properties of old and new techniques in detrended analysis of time series, Acta Phys. Polon. B 36 (2005) 2403–2413.
[21] J. Barunik, L. Kristoufek, On Hurst exponent estimation under heavy-tailed distributions, Physica A 389 (18) (2010) 3844–3855.
[22] F. Serinaldi, Use and misuse of some hurst parameter estimators applied to stationary and non-stationary financial time series, Physica A 389 (2010) 2770–2781.
[23] G.-F. Gu, W.-X. Zhou, Detrending moving average algorithm for multifractals, Phys. Rev. E 82 (2010) 011136.
[24] A. Schumann, J. Kantelhardt, Multifractal moving average analysis and test of multifractal model with tuned correlations, Physica A 390 (2011) 2637–2654.
[25] Z.-Q. Jiang, W.-X. Zhou, Multifractal detrending moving average cross-correlation analysis, Phys. Rev. E 84 (2011) 016106.
[26] Y.-H. Shao, G.-F. Gu, Z.-Q. Jiang, D. Zhou, W.-X. an Sornette, Comparing the performance of FA, DFA and DMA using different synthetic long-range correlated time series, Sci. Rep. 2 (2012) 835.