The publication is available at https://doi.org/10.1007/978-3-319-69989-9_24. Cite as:
M. Janowicz, L. J. Chmielewski, J. Kaleta, L. Ochnio, A. Orłowski, and A. Zembrzuski. Persistent Correlations in Major Indices of the World Stock Markets. In C. Berger-Vachon, A. M. Gil-Lafuente, J. Kacprzyk, Y. Kondratenko, J. M. Merigó and F. C. Morabito, editors, Complex Systems: Solutions and Challenges in Economics, Management and Engineering. Dedicated to Professor Jaime Gil Aluja, volume 125 of Studies in Systems, Decision and Control, pages 411-422. Springer, Cham, 2018. doi:10.1007/978-3-319-69989-9_24.

Chapter 24
Persistent Correlations in Major Indices of the World Stock Markets

Maciej Janowicz, Leszek J Chmielewski, Joanna Kaleta, Luiza Ochnio,
Arkadiusz Orłowski and Andrzej Zembrzuski

Abstract Time-dependent cross-correlation functions have been calculated between returns of the major indices of the world stock markets. One-, two-, and three-day shifts have been considered. Surprisingly high and persistent-in-time correlations have been found among some of the indices. Part of those correlations can attributed to the geographical factors, for instance, strong correlations between two major Japanese indices have been observed. The reason for other, somewhat exotic correlations, appear to be as much accidental as it is apparent. It seems that the observed correlations may be of practical value in the stock market speculations.

Key words: stock market indices, correlation functions, Pearson correlation, technical analysis

24.1 Introduction

As is well known, the time-evolution of the stock markets, and, in particular, the stock market indices, exhibit considerable short-time correlations. These are sufficient to disprove claims of the reliability of the random-walk approximations to

Maciej Janowicz
e-mail: maciej_janowicz@sggw.pl

Leszek J Chmielewski
e-mail: leszek_chmielewski@sggw.pl

Joanna Kaleta
e-mail: joanna_kaleta@sggw.pl

Luiza Ochnio
e-mail: luiza_ochnio@sggw.pl

Arkadiusz Orłowski
e-mail: arkadiusz_orlowski@sggw.pl

Andrzej Zembrzuski
e-mail: andrzej_zembrzuski@sggw.pl

Faculty of Applied Informatics and Mathematics – WZIM
Warsaw University of Life Sciences – SGGW
Nowoursynowska 159, bldg 34
02-776 Warsaw, Poland
www.wzim.sggw.pl

that evolution or strong versions of the efficient-market hypothesis [1, 2]. On the other hand, there exist a set of usually quite simple computational and visualization techniques, called technical analysis [3, 4], which aims to obtain the approximate predictions of trends and their corrections in market the data. It is so even though the data appear as realizations of a *random* process. It is to be noticed that some recent publications, e.g. [5, 6, 7, 8] have lead to considerable revision of the previously ultra-critical stand of the many experts regarding technical analysis.

One of the many possible strategies of "beating the market" which are close to the spirit of technical analysis consists of identification of correlated pairs of stocks. In fact, if we find that, during a specific, long, time interval, the increase (decrease) of the value of one stocks has been followed by a corresponding change in another stock, we may attempt to guess that this relation may persist, at least for short additional time.

To quantify the above intuitive remarks, we have computed the time-dependent cross-correlation functions of returns of the major world stock market indices. Among the latter, the following indices have been included: ALL ORDINARIES (ALL-ORD), AMEX MAJOR (AMEX-MAJ), BOVESPA, B-SHARES, BUENOS, BUX, CAC40, DAX, DJIA, DJTA, DJUA, EOE, FTSE100, HANGSENG, MEXICIPS, NASDAQ, NIKKEI, RUSSELL, SASESLCT, SMI, SP500, TOPIX, and TSE300. Let us notice that in one of our previous work a similar research but involving *normalized values* has been reported. Here, however, we deal with correlations of daily *returns*.

The main body of this work is organized as follows. In Section 2 we describe our procedure; rhe description as it is, in fact, very simple, and involves well-known quantities. Section 3 is devoted to the presentation of results. Finally, Section 4 comprises some concluding remarks.

24.2 Correlation Functions of World Stock Market Indices

Let $K_n^{(a)}$ denote the closing value at the trading day n of a stock market index (a), and let $Z_n^{(a)}$ denote the corresponding relative return, i.e. $Z_n^{(a)} = (K_n^{(a)} - K_{n-1}^{(a)})/K_{n-1}^{(a)}$. We define the cross-correlation function of returns of two indices (a) and (b) as:

$$C^{(a,b)}(n,M,t) = \frac{Cov\left(Z_{n-M}^{(a)} Z_n^{(b)}\right)}{\sqrt{Var\left(Z_{n-M}^{(a)}\right) Var\left(Z_n^{(b)}\right)}}, \tag{24.1}$$

where t denotes the time interval over which the averaging in the calculations of the variances and covariance has been performed (unfortunately, we have no ensemble-averaging in our disposal here).

Thus, we have defined the correlation function in terms of the Pearson R coefficient [9] made of two sequences of the length t. A similar definition can be given in terms of, e.g., Spearman coefficient.

It is to be noted that $C^{(a,b)}(n,M,t)$ is a function of three time variables. In the following, we have considered only three values of M, $M = 1,2,3$. The averaging time t has been varied from $t = 60$ to $t = N_0$, with N_0 being the largest value of the trading sessions common to all indices. N_0 has been equal to 3452 enclosing the time interval from 10th of October, 2001 to 31st of October, 2016). The data for stock market indices have been downloaded from [11]. The correlation functions have been computed using `pearsonr` function from the Python (2.7) module *scipy.stats*, version 0.14.0 [10].

24.3 Results and Discussion

We have started our study with n equal to t. Firstly, we have considered Pearson's R coefficient using all available data.

In Tables 24.1-24.3 we have provided pairs of indices with largest values of the correlation functions for $n = t = u$, $u = N_0 - 4$, for $M = 1, 2, 3$.

Table 24.1 List of indices with largest correlation functions for $n = t = N_0 - 4$, $M = 1$. In all cases p-value has been smaller than 0.05.

Index (a)	Index (b)	$C^{(a,b)}(u,1,u)$
NIKKEI	TOPIX	0.785
NASDAQ	SP500	0.694
NASDAQ	RUSSELL	0.690
NASDAQ	DJTA	0.651
DJIA	NASDAQ	0.474
DJIA	RUSSELL	0.411

Even though the number of cross-correlation functions with the values greater than 0.4, say, has been rather small, in particular cases those correlations have been very significant for $M = 1$ as specified in Table 24.1. One might say, for instance, that during several years one can predict the behavior of the TOPIX index from the knowledge of the NIKKEI 225 index change one day before. Remarkably, it has worked only one way. What is more, for t significantly smaller than u a quite interesting structure in the n-dependence of the correlation function occurs, please see below.

Table 24.2 List of indices with largest correlation functions for $n = t = N_0 - 4$, $M = 2$. In all cases p-value has been smaller than 0.05.

Index (a)	Index (b)	$C^{(a,b)}(u,2,u)$
DJIA	SP500	0.388
DJIA	DJUA	0.366
DJIA	DJTA	0.335
DJIA	RUSSELL	0.325
DJIA	MEXICIPC	0.256

The above relatively high values of the correlation functions between the returns of DJIA and the returns of other American (and Mexican) indices certainly cannot be called counterintuitive. However, it is quite remarkable that the pairs listed in Table 24.2 do *not* appear in the list of highly correlated pairs for the shift $M \neq 2$. Please see also further comments following Fig. 24.2.

Table 24.3 List of indices with largest correlation functions for $n = t = N_0 - 4$, $M = 3$. In all cases p-value has been smaller than 0.05.

Index (a)	Index (b)	$C^{(a,b)}(u,3,u)$
SMI	TSE-300	0.159
FTSE100	ALL-ORD	0.135
DJIA	MEXICIPC	0.105
EOE	SMI	0.092
SP500	TSE-300	0.091
AMEX-MAJ	TSE-300	0.090

For $M = 3$ we have not observed any large values of the R coefficient, for $M > 3$ they become even smaller. Let us also report that we have not obtained any negative R coefficient with significant absolute value for any M. The indices definitely tend to be correlated rather than anti-correlated.

In the Tables 24.4-24.6 we have listed the same pairs of indices as in Tables 24.1-24.3 but accompanied with the numbers of correct and incorrect predictions. Our definition of a correct prediction is trivial: if the sign of the product of returns of index (a) at time t and index (b) at time $t+M$ is positive, we say that the prediction is correct; otherwise the prediction is incorrect.

Table 24.4 List of indices with largest correlation functions for $n = t = N_0 - 4$, $M = 1$ with the numbers of correct and incorrect predictions of the behavior one the second-column indices based on the change of the corresponding first-column ones.

Index (a)	Index (b)	No. of correct preditions	No. of incorrect predictions
NIKKEI	TOPIX	2852	596
NASDAQ	SP500	2296	1089
NASDAQ	RUSSELL	2368	1080
NASDAQ	DJTA	2301	1147
DJIA	NASDAQ	2359	1089
DJIA	RUSSELL	2281	1167

Table 24.5 List of indices with largest correlation functions for $n = t = N_0 - 4$, $M = 1$ with the numbers of correct and incorrect predictions of the behavior one the second-column indices based on the change of the corresponding first-column ones.

Index (a)	Index (b)	No. of correct preditions	No. of incorrect predictions
DJIA	SP500	2086	1362
DJIA	DJUA	1953	1495
DJIA	DJTA	2042	1406
DJIA	RUSSELL	1953	1495
DJIA	MEXICIPC	1887	1561

The content of the Tables 24.4-24.6 largely agrees with the conclusions one can draw from three previous tables. For the first five rows of Table 24.4 we can realize that the number of correct predictions is at least two times larger that the incorrect ones. Quite obviously, even such a ratio could not guarantee any successful trading on any index-based financial instruments. On the other hand, even the smallest ad-

Table 24.6 List of indices with largest correlation functions for $n = t = N_0 - 4$, $M = 3$ with the numbers of correct and incorrect predictions of the behavior one the second-column indices based on the change of the corresponding first-column ones.

Index (a)	Index (b)	No. of correct preditions	No. of incorrect predictions
SMI	TSE-300	1833	1516
FTSE100	ALL-ORD	1790	1658
DJIA	MEXICIPC	1828	1620
EOE	SMI	1777	1671
SP500	TSE-300	1885	1563
AMEX-MAJ	TSE-300	1813	1635

vantage like that provided by Table24.6 may sometimes be sufficient to get the edge on trading competitors.

Let us notice that the difference between correct and incorrect predictions in the case of EOE-SMI pair is so small that p-value in the G-test is ≈ 0.07. Thus, one cannot exclude the zeroth hypothesis that the difference between the numbers of predictions if purely accidental. In all other cases we have had p-value in the G-test smaller than 0.05.

We have also performed calculations for the values of n and t different from u. For instance, we have obtained the R coefficient from the last 480 trading sessions (approximately two years) preceding 31st of October, 2016. For the pair CAC40 - EOE a record value of R equal to 0.94 has been obtained with the percent of correct predictions equal to 0.875. For $M = 2$ the most considerable correlation has been obtained between the DAX and SMI indices; R has been equal to 0.526, and the ratio of correct to incorrect predictions has almost exactly been equal to $2:1$.

We have not, however, been satisfied with the above results for it has not been clear precisely how the correlation function depends on the time n if the averaging window t is smaller than n.

In the following figures we have displayed selected cross-correlation function for given M and t as functions of n. These have been compared with the time evolution of the indices themselves. Note that it is the closing values of the index (not the returns) which have been shown in the central and lower parts of each figure.

Naturally, the graphs in Fig. 24.1 provide additional insight into the results tabulated in Tables 24.1 and 24.4. In fact, we can see that correlation function involving NIKKEI and TOPIX has not been constant, and there have been two time intervals in which the correlation dropped quite significantly. The second one includes recent trading sessions. As can be seen from the central and lower parts of Figure 24.1, there is nothing in the values of indices themselves which could correspond to the drops of correlations. The first period of low correlations may, perhaps, be attributed to the change of the weighting system with the help of which the TOPIX index has been calculated; that change was completed in January 2006. We have failed, however, to identify a possible reason of breaking down the correlation in the beginning of 2015.

The history of correlations of the indices DJIA and SP500 ($M = 2$), which, perhaps, might be considered as the most important and most obvious of all, is even more striking. It appears that the quite considerable value of R coefficient calculated during the last 14 years was built during the relatively short time interval 2006-2010. Recently, DJIA-SP500 have become almost completely “unreliable”, so to say.

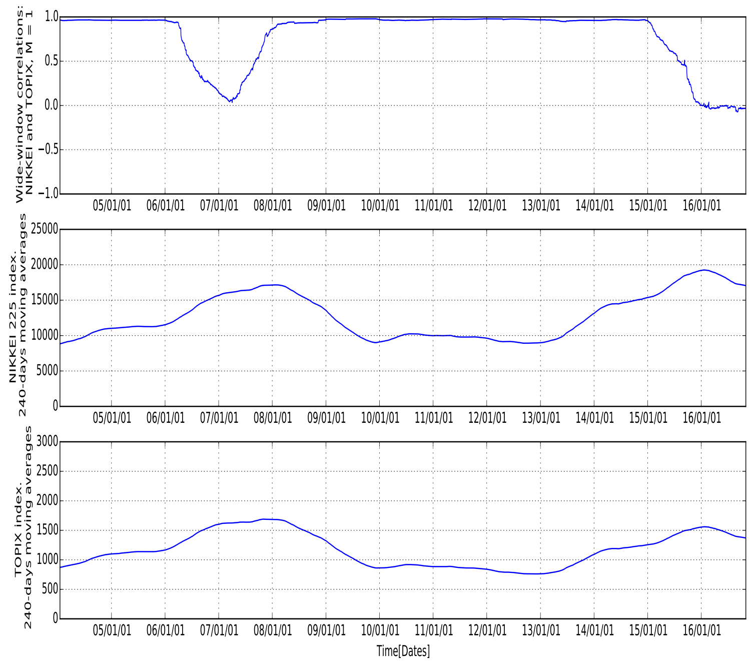

Fig. 24.1 Upper part: time-dependent correlation function of returns for NIKKEI 225 and TOPIX indices as a function of discrete time n for $M = 1$, $t = 240$; central part: simple moving average (with window 240) of closing values of the NIKKEI 225 index in the same time interval; lower part: corresponding values of the TOPIX index.

We have provided Figures 24.4-24.6 largely to confirm the results displayed in the three previous figures. One can convince oneself that the shorter intervals and smaller windows of moving averages do not not change the overall picture and provide, on this occasion, a little more than noise.

24.4 Concluding Remarks

In this work we have computed cross-correlation function of time series generated by the returns of of important world stock market indices. We have identified pairs of indices such that correlations are both strong and persistent in time. Preliminary assessments of predictive power of the correlations have been performed. We believe that, with sufficient care, the knowledge of correlations between indices and individual stocks may be used in practice, possibly even to enhance the predictive power of technical indicators for trading purposes. Needless to say, extreme caution is required in any such attempts. In fact, they can break down at any time. For instance, when we used historical data up to the middle summer of 2016, we have observed the highest correlation function for $M = 3$, $n = t = u - 3$ in the pair MEXICIPC - SMI. During the subsequent mounts that correlation deteriorated quite

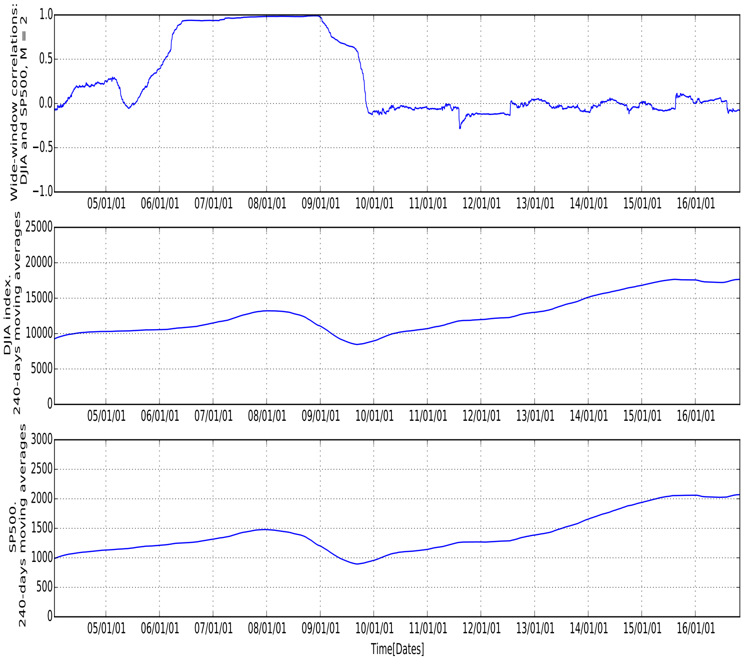

Fig. 24.2 Upper part: time-dependent correlation function of returns for DJIA and SP500 indices as a function of discrete time n for $M = 2$, $t = 240$; central part: simple moving average (with window 240) of closing values of the DJIA index in the same time interval; lower part: corresponding values of the SP500 index.

spectacularly. One might think that it would be a valuable enterprise to find any indicators which could suggest that the correlations have just started to build up or are just about to get ruined, provided that such indicators exist. Finally, one may reasonably argue that the above very simple correlation analysis can and should be combined with the analysis of cointegration to have deeper insight into connection between shifted (i.e. with $M \neq 0$) and unshifted time series.

References

1. Malkiel, B.: A Random Walk Down the Wall Street, Norton, New York (1981).
2. Fama, E., Blume, M.: E.F. Fama and M. Blume filter rules and stock-market trading, Journal of Business **39** (1966) 226–241.
3. Murphy, J.: Technical Analysis of Financial Markets, New York Institute of Finance, New York (1999).
4. Kaufman, P.: Trading Systems and Methods, John Wiley and Sons, New York (2013).
5. Brock, W., Lakonishok, J., LeBaron, B.: Simple technical trading rules and the stochastic properties of stock returns, Journal of Finance **47**(5) (1992) 1731–1764.
6. Lo, A., MacKinley, A.: Stock market prices do not follow random walks: Evidence from a simple specification test, Review of Financial Studies **1** (1988) 41–66.
7. Lo, A., MacKinley, A.: A Non-Random Walk down Wall Street, Princeton University Press, Princeton (1999).

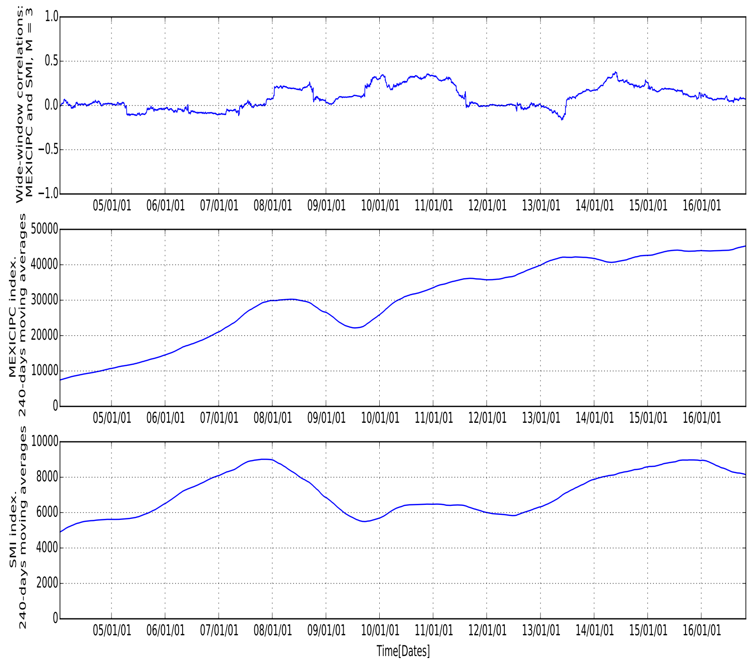

Fig. 24.3 Upper part: time-dependent correlation function of returns for MEXICIPC and SMI indices as a function of discrete time n for $M = 3$, $t = 240$; central part: simple moving average (with window 240) of closing values of the MEXICIPC index in the same time interval; lower part: corresponding values of the SMI index.

8. Lo, A., Mamaysky, H., Wang, J.: Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation, Journal of Finance **55**(4) (2000) 1705–1765.
9. Pearson, K.: Notes on regression and inheritance in the case of two parents, Proceedings of the Royal Society of London **58** (1895) 240–242.
10. Scipy Community: Statistical functions in Python (2016) https://docs.scipy.org/doc/scipy/reference/stats.html.
11. Bossa.pl, http://bossa.pl/notowania/metastock [Accessed 2016-Nov-15]

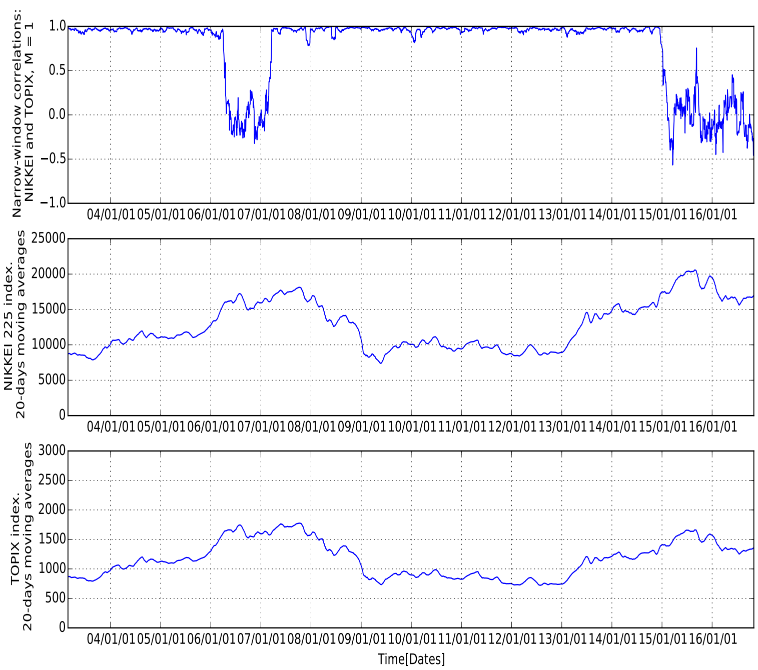

Fig. 24.4 The same as in Figure 24.1 but for $t = 60$ and moving averages with the window length equal to 20.

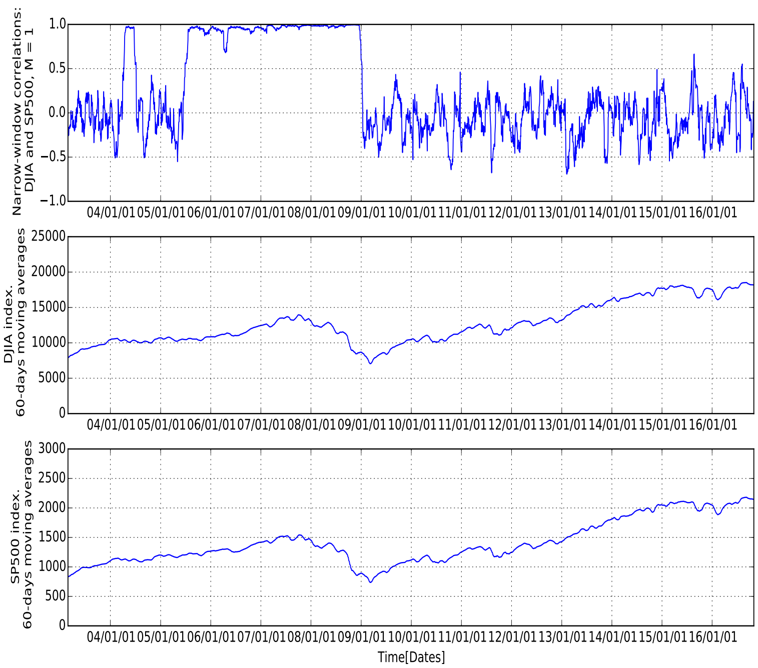

Fig. 24.5 The same as in Figure 24.2 but for $t = 60$ and moving averages with the window length equal to 20.

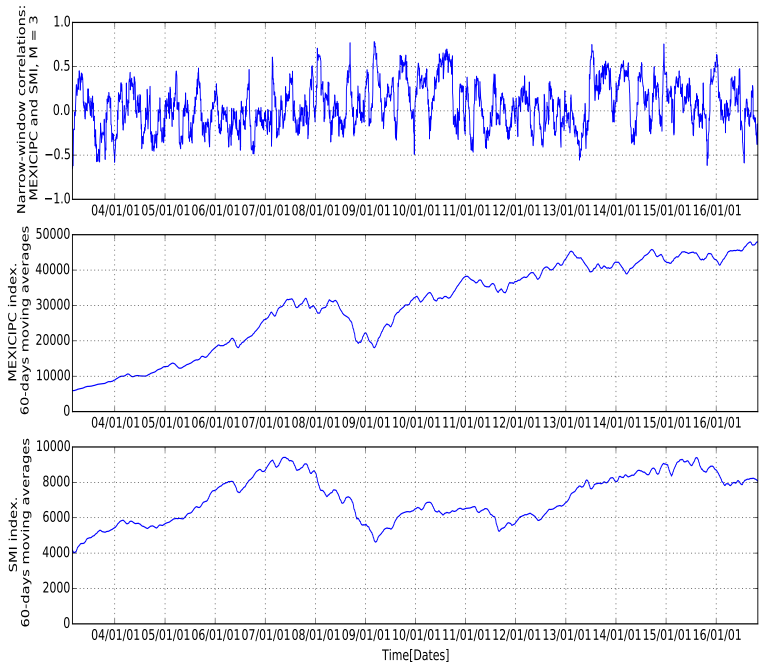

Fig. 24.6 The same as in Figure 24.3 but for $t = 60$ and moving averages with the window length equal to 20.