MPRA

Munich Personal RePEc Archive

Analysis of the Lead-Lag Relationship on South Africa capital market

Marcel Rešovský and Marek Gróf and Denis Horváth and Vladimír Gazda

Technical University of Košice, Faculty of Economics, Pavol Jozef Šafárik University in Košice, Institute of Physics

2014

Analysis of the Lead-Lag Relationship on South Africa capital market

Marcel Rešovský, Marek Gróf, Denis Horváth, Vladimír Gazda

Abstract

Despite the efficient market hypothesis (EHM), lead-lag relationships can be observed mainly between financial derivatives and underlying asset prices, prices of large and small companies, etc. In our paper, we examined the lead-lag relationship between prices of open ended funds and an all-share index as a representative of the capital market. Along with more traditional methods of using cross correlations, partial correlation and Toda-Yamamoto causality tests, we also analysed the speed of adjustment of assets to their intrinsic values and identified the most prevalent lag using rolling time windows. The analysis was performed using data from the South Africa capital market.

Keywords: lead-lag relationship, open ended fund, all-share index, causality, efficient market hypothesis

1 Introduction

The efficient market hypothesis (EMH) belongs without dispute to one of the most influential theories in the field of financial economics. Its main criticism stems from its assumptions that are considered not to be very realistic (see Fama, 1970, p. 387). Roughly speaking, it is not a theory, which can be accepted as a strict scientific truth because it is not empirically falsifiable nor verifiable. Despite this fact, a lot of attempts have been made to find some supportive arguments for the theory, all the while counter-examples have been presented in order to evade or devalue the theory. On the other hand, a wide variety of methodological tools and frequently used techniques have been inspired by the EMH idea, stating that all available information clears the market and the current price of the underlying assets is also the best predictor of its future price. Repeated observations (calendar anomalies – French, 1980; Lakonishok and Smidt, 1988; etc.; size effects – Keim, 1983 represent just some of the introductory seminal work boosting research in the end of the 20th century) contradicted this idea and became the subject of extensive research. On the other hand, appearance of these effects appealed to investors, allowing them to gain extra profits, which again cleared the market and quickly eliminated the arbitrage opportunities (Malkiel, 2003; Basher and Sadorsky, 2006

are some good examples of the research in this domain). The contradictory discussion could be interpreted as a demonstration of the Goodhart's law in financial economics. Discovering new examples by extending the spectrum of non-trivial market price predictors, empirical research has in part falsified the EMH. Among the more recent examples is the so called *lead-lag relationship*, bringing the evidence of market frictions, where prices of particular assets can lead (or lag behind) other ones. If we focus solely on the stock markets, we can mention the work of Conover and Peterson (1999), in which the effect between option and stock prices is studied. Similarly, the work of De Medeiros et al. (2009) describes the lead-lag relationship between the U.S. and Brazilian stock markets, Lo and MacKinlay (1990) and Hou (2007) presented a study of large firms stock's returns leading the market. In our work, we focus our attention solely on the existence of the lead-lag relationships between price returns of 8 open ended funds and logarithmic price changes of an all-share index on the South Africa capital market (it is not the ambition of this paper to provide further policy or strategy implications of this relationship).

The paper is organised as follows. In Section 2 we present the collection of data records and we summarise the primary statistical facts of the time-series we plan to manipulate. In Section 3 we present some empirical facts supporting the hypothesised lead-lag relationship between price returns of open ended funds and logarithmic price changes of the all-share index, which implies deviation from the efficient market hypothesis. In Section 4, vector autoregression is applied to test the lead-lag causality. Section 5 presents new results regarding the application of the model of intrinsic value. The rolling-window approach to the lead-lag relationship is discussed in Section 6.

2 Data description and summary statistics

The data used has been obtained from the Bloomberg database. It contains daily time series of the prices of eight selected open ended funds (further denoted as Funds) on the South Africa capital market and daily times series of prices of FTSE/JSE All Share Index (denoted as Index) from January 2, 2009 to November 15, 2013. The studied assets, time period and market it self were chosen at random. The time series consisted of 1221 observations. The real time progress of Fund prices and prices of the Index is depicted in Figure 1. The descriptive statistics of the corresponding log-returns

$$r_{i,t}^{\mathrm{F}} = \ln\left(\frac{P_{i,t}^{\mathrm{F}}}{P_{i,t-1}^{\mathrm{F}}}\right) \qquad r_t^{\mathrm{I}} = \ln\left(\frac{P_t^{\mathrm{I}}}{P_{t-1}^{\mathrm{I}}}\right) \tag{2.1}$$

is shown in Table 1.

Inspecting the statistics we came to the conclusion that the mean returns of Funds do not predominate over the returns of the Index and vice versa, however in two cases Funds present higher mean returns than the Index. In addition, the standard deviations of the returns of five Funds are smaller than the one for the Index. If

	Index	F1	F2	F3	F4	F5	F6	F7	F8
Mean	0.06	0.02	0.06	0.06	0.06	0.06	0.06	0.02	0.09
Median	0.09	0.03	0.10	0.08	0.08	0.10	0.09	0.06	0.12
1. Quar.	-0.57	-0.82	-0.55	-0.48	-0.42	-0.63	-0.52	-0.29	-0.33
3. Quar.	0.70	1.02	0.72	0.66	0.59	0.80	0.71	0.38	0.56
St. dev.	1.12	1.47	1.13	0.93	0.87	1.22	1.03	0.65	0.77
Skew.	-0.01	-0.22	-0.20	-0.22	-0.19	-0.14	-0.23	-0.40	-0.26
Kurt.	1.46	0.81	1.92	1.26	1.76	1.40	1.26	4.37	1.21
AR(1)	0.02	-0.01	-0.03	0.03	0.02	-0.01	0.03	0.04	0.07

Table 1: Descriptive statistics of the log-price returns. Comparison of the Funds and Index. The values are multiplied by 100.

we look at the time-dynamic behaviour using the autocorrelation coefficients, we see they are rather small which implies a low level of short-term memory.

3 Lead-Lag relationship – analysis

In this section we would like to answer the question whether there are some significant lead-lag relations between the returns of Funds and log-price changes of the Index. To demonstrate it, we follow the methodology of Chiao et al. (2004) and Chordia and Swaminathan (2000). Here, we tackle estimation of the cross correlations between the log-price changes of Index and lagged/led returns of Funds. The results presented as histograms are shown in Figure 2. We see that the cross-correlations of the returns of Funds lagged by one period are significant for all Funds. Similar results have been reported by Chordia and Swaminathan (2000), who inspected the lead-lag relationship between size based portfolios. As the figure shows, spikes at $lag = 1$ dominate over the spikes at other lags. Now, the question whether the significant correlation is just a demonstration of relationship between the current Index and Fund returns and short-term memory of the Index log-price changes becomes pertinent. Although, as it is seen in Table 1, the short-term memory seems not to play a significant role. This preliminary finding stimulated our interest to explore partial correlations (pcor) listed in Table 2. What is quite striking is the dominance

$$\text{pcor}(r_{i,t}^{\text{F}}, r_{t-1}^{\text{I}}; r_t^{\text{I}}) > \text{cor}(r_{i,t}^{\text{F}}, r_{t-1}^{\text{I}}) \tag{3.1}$$

valid for all the Index-Fund pairs studied. In all the cases the equal time cross-correlation $\text{cor}(r_{i,t}^{\text{F}}, r_t^{\text{I}})$ dominated over both aforementioned $lag = 1$ correlations. In addition, our suspicion of lead-lag effects is supported by the fact that all Fund-Index cross-correlations far exceed the autocorrelation of the Index $\text{cor}(r_t^{\text{I}}, r_{t-1}^{\text{I}})$.

	F1	F2	F3	F4	F5	F6	F7	F8
pcor $(r_{i,t}^{\mathrm{F}}, r_{t-1}^{\mathrm{I}}; r_t^{\mathrm{I}})$	0.22	0.25	0.33	0.34	0.32	0.35	0.26	0.32
cor$(r_{i,t}^{\mathrm{F}}, r_t^{\mathrm{I}})$	0.78	0.64	0.75	0.75	0.83	0.80	0.30	0.66
cor$(r_t^{\mathrm{I}}, r_{t-1}^{\mathrm{I}})$	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
cor$(r_{i,t}^{\mathrm{F}}, r_{t-1}^{\mathrm{I}})$	0.15	0.20	0.23	0.24	0.19	0.23	0.26	0.25

Table 2: Table of correlation coefficients of funds F1,..., F8 and Index at specified lags ($lag = 0$ and $lag = 1$). The first table line contains partial correlations (pcor) measuring the strength of association of $r_{i,t}^{\mathrm{F}}$, (returns of Funds $i = 1, 2, \ldots, 8$) and r_{t-1}^{I} (log-price changes of Index) with eliminated effect of the additional variable r_t^{I}. The next three lines present specific ordinary correlation coefficients (cor) of interest. Partial correlations are significant at the level of significance $\alpha = 0.05$.

4 Vector autoregression and Toda-Yamamoto causality tests

The correlation analysis presented in Table 2 is based on pairwise relations and is not capable of capturing higher order lag dynamics. Therefore, we decided to use the Vector Autoregression (VAR) methodology followed by the Toda-Yamamoto (see Toda and Yamamoto, 1995) causality test. For the Index and i-th Fund the system of regression equations takes the form

$$\ln(P_t^{\mathrm{I}}) = \sum_{\tau=1}^{p+dMax} \alpha_\tau \ln(P_{t-\tau}^{\mathrm{I}}) + \sum_{\tau=1}^{p+dMax} \beta_{i,\tau} \ln(P_{i,t-\tau}^{\mathrm{F}}) + \epsilon_t \,, \qquad (4.1)$$

$$\ln(P_{i,t}^{\mathrm{F}}) = \sum_{\tau=1}^{p+dMax} \gamma_{i,\tau} \ln(P_{i,t-\tau}^{\mathrm{F}}) + \sum_{\tau=1}^{p+dMax} \delta_\tau \ln(P_{t-\tau}^{\mathrm{I}}) + \nu_{i,t} \,, \qquad (4.2)$$

where p is an optimal VAR lag selected by the Schwarz Information Criterion, $dMax$ is the highest order of integration in the $\ln(P_t^{\mathrm{I}})$ and $\ln(P_{i,t}^{\mathrm{F}})$ pair and ϵ_t, ν_t are the error terms. The reason why we apply Toda-Yamamoto instead of the more frequently used Granger causality was that the Johansen cointegration tests needed to perform the Granger causality tests may give fairly ambiguous results (high sensitivity to the lags included, significance levels near the indecisive 10% levels, contradictory results to the evidence of the Dickey Fuller cointegration tests). On the other hand, there is well established Toda-Yamamoto methodology, which provides less restrictive causality test with the following properties:

1. Granger test gives misleading results in the case of faulty identified cointegration among the variables.

2. Granger causality tests in Error Correction Model (ECM) still contain the possibility of incorrect inference and suffer from the asymptotic nuisance parameter in some cases (see Toda and Phillips, 1993, p. 1338).

3. Toda-Yamamoto test does not require knowledge of the integration and cointegration properties of the data. It can be applied even when the underlying time series are not of the same order, and when the rank conditions fail (Toda and Yamamoto, 1995, p. 225).

First, we tested the integration of the time series by the traditional Augmented Dickey-Fuller (ADF) and most widely used Kwiatkowski-Phillips-Schmitt-Shin (KPSS) tests, which formulate the zero hypotheses in a complementary way (presence of unit root vs. stationarity). The benefit of such simultaneous testing is that it reduces the risk of decision errors. In all investigated time series the alternative ADF test hypothesis on the non-existence of the unit root on levels was not rejected at any standard significance level, while in the case of the KPSS test the zero stationarity hypothesis has been rejected at 1% level for each time series. After differencing, testing by the means of ADF tests yields acceptance of the alternative no unit root hypothesis for each time series at the 1% level.

Considering all the investigated time series data on prices to be of order $I(1)$, we run the regression of Eq. (4.1) and Eq. (4.2). The Toda-Yamamoto procedure uses modified Wald statistics for testing of

$$\beta_1 = \beta_2 = \cdots = \beta_p = 0 \tag{4.3}$$

in case of $F \Rightarrow I$ causality and

$$\delta_1 = \delta_2 = \cdots = \delta_p = 0 \tag{4.4}$$

for the opposite causal association $I \Rightarrow F$. The test results are given in Table 3. The relatively high number of observations (1221) helps us minimise the finite sample bias and ensures a good approximation of F-Wald statistics, which follows a chi-squared distribution with p degrees of freedom.

	F1	F2	F3	F4
p	2	2	2	2
d	1	1	1	1
$W(I \Rightarrow F)$	41.01*	56.3*	71.23*	79.36*
$W(F \Rightarrow I)$	0.60	3.42	1.54	1.99
p	6	3	2	2
d	1	1	1	1
$W(I \Rightarrow F)$	39.4***	58.52**	47.54*	47.29*
$W(F \Rightarrow I)$	1.62	2.35	0.88	1.62

Note: Signif. codes: ()*** 0.001; ()** 0.01; ()* 0.05; (). 0.1

Table 3: Toda-Yamamoto tests supporting the $I \Rightarrow F$ causality.

5 Lead-lag relationship as a result of different adjustments to the intrinsic values

Lead-lag relationship is often estimated by the Vector autoregression. In our trial implementation of this approach, the error term was found to follow a persistent systematic pattern observable predominantly at higher levels of autocorrelations as well as statistically significant cross correlations. That is why we decided to apply another methodology. Our choice was guided by the work of Theobald and Yallup (1998). In the following we present an adaptation of this methodology and axiomatics to our specific situation. The original single equation model is now decomposed into the system of eight plus one stochastic equations for the partial adjustments

$$\ln P_{i,t}^{\mathrm{F}} - \ln P_{i,t-1}^{\mathrm{F}} = g_i^{\mathrm{F}} \left(\ln V_{i,t}^{\mathrm{F}} - \ln P_{i,t-1}^{\mathrm{F}} \right) + u_{i,t}^{\mathrm{F}}, \qquad i = 1, 2, \ldots, 8, \quad (5.1)$$
$$\ln P_t^{\mathrm{I}} - \ln P_{t-1}^{\mathrm{I}} = g^{\mathrm{I}} \left(\ln V_t^{\mathrm{I}} - \ln P_{t-1}^{\mathrm{I}} \right) + u_t^{\mathrm{I}},$$

where $V_{i,t}^{\mathrm{F}}$ and V_t^{I} are the intrinsic values of the i-th Fund and Index at time t; $g_i^{\mathrm{F}} \in [0; 2]$, $g^{\mathrm{I}} \in [0; 2]$ are the speeds of adjustment to the corresponding intrinsic values $V_{i,t}^{\mathrm{F}}$, V_t^{I}. The main assumption here is that the intrinsic values are assumed to be the logarithmic random walks

$$\ln V_{i,t}^{\mathrm{F}} = \ln V_{i,t-1}^{\mathrm{F}} + v_{i,t}^{\mathrm{F}}, \qquad \ln V_t^{\mathrm{I}} = \ln V_{t-1}^{\mathrm{I}} + v_t^{\mathrm{I}}, \qquad (5.2)$$

where $u_{i,t}^{\mathrm{F}}$ and u_t^{I} are the i.i.d. zero mean white noise terms. Two distinct types of the behaviour should be mentioned: if, for example, $g_i^{\mathrm{F}} = 0$ then the asset prices do not reflect their intrinsic values and form the random walk, whereas in the idealised case $g_i^{\mathrm{F}} = 1$, the Fund prices fully reflect their intrinsic values, which are the only determinants of their expected values. The interpretation of g^{I} is analogous. The equations of the system Eq. (5.1) are coupled to each other due to postulated cross-covariances

$$\mathrm{cov}(u_{i,t-t'}^{\mathrm{F}}, v_{j,t-t''}^{\mathrm{F}}) = 0 \quad \text{for all} \quad i, j, t', t'', \qquad (5.3)$$
$$\mathrm{cov}(u_{t-t'}^{\mathrm{I}}, v_{t-t''}^{\mathrm{I}}) = 0 \quad \text{for all} \quad t', t'',$$
$$\mathrm{cov}(u_{i,t-t'}^{\mathrm{F}}, v_{t-t''}^{\mathrm{I}}) = 0 \quad \text{for all} \quad i, t', t'',$$
$$\mathrm{cov}(u_{t-t'}^{\mathrm{I}}, v_{j,t-t''}^{\mathrm{F}}) = 0 \quad \text{for all} \quad j, t', t'',$$
$$\mathrm{cov}(u_{t-t'}^{\mathrm{I}}, u_{j,t-t''}^{\mathrm{F}}) = 0 \quad \text{for all} \quad j, t', t'',$$
$$\mathrm{cov}(v_{t-t'}^{\mathrm{I}}, v_{j,t-t''}^{\mathrm{F}}) = 0 \quad \text{for all} \quad j, t' \neq t'',$$
$$\mathrm{cov}(v_{t-t'}^{\mathrm{I}}, v_{j,t-t'}^{\mathrm{F}}) = \rho_j^{\mathrm{I,F}} \quad \text{for all} \quad j, t',$$

where ρ_j^{IF} is the amplitude of the equal time cross-covariances of $v_{t-t'}^{\mathrm{I}}$ and $v_{j,t-t'}^{\mathrm{F}}$. Its numerical value is irrelevant for the estimation of the speeds of the adjustments

$$g^{\mathrm{I}} = 1 - \frac{\mathrm{cov}(r_{i,t-1}^{\mathrm{F}}, r_t^{\mathrm{I}})}{\mathrm{cov}(r_{i,t}^{\mathrm{F}}, r_t^{\mathrm{I}})} \qquad (5.4)$$
$$g_i^{\mathrm{F}} = 1 - \frac{\mathrm{cov}(r_{t-1}^{\mathrm{I}}, r_{i,t}^{\mathrm{F}})}{\mathrm{cov}(r_t^{\mathrm{I}}, r_{i,t}^{\mathrm{F}})}$$

The results of the calculation are depicted in Tab.4. The value $g^{\mathrm{I}} > 1$ indicates that Index overreacts to its intrinsic values. The effect of the lead-lag relationship is empirically assessed by showing that the inequality $g^{\mathrm{I}} > g_i^{\mathrm{F}}$ is satisfied for all Funds.

	F1	F2	F3	F4	F5	F6	F7	F8
g_i^{F}	0.81	0.68	0.69	0.68	0.77	0.71	0.15	0.62

Table 4: The differences in the speeds of adjustment calculated for Funds by means of Eq. (5.4). Moreover, the overreaction $g^{\mathrm{I}} \sim 1.02$ is obtained for the Index.

6 Temporal variations and local structural analysis of data

The previous results have been based on using of the entire data sample. In the present section, we focus more on the local subset analysis, which admits to shed light on the inhomogeneous structure of data and local processes. The careful reader may recognise such attempt as application motivated by resembling the real time conditions, where a time-window is moving down the data and emphasis is placed on the role of instant feedback since future data is temporarily unavailable. Each correlation is considered as moving average on the rolling window. Thus for each time t the maximum of the correlation function estimate may serve to define the instant optimal two-element tuples $(i^*, lag^*)_t$ as follows

$$(i^*, lag^*)_t \;=\; \arg\max_{\substack{lag \in \{1, \ldots, 20\} \\ i, \ldots, 8}} \underset{\substack{\text{estimated} \\ \text{on interval } [t-120, t]}}{\mathrm{cor}(r_{i,\tau}^{\mathrm{F}}, r_{\tau-lag}^{\mathrm{I}})} \tag{6.1}$$

where i runs over the indices of the Funds. More precisely, the choice of lag^* (and thus i^*) is suboptimal since it is constrained by the lower bound 1 and upper bound 20 days (in part because of computational reasons). This task is specific due to the fact that right from the start we are penalising the situations with negative and zero lag^*. We analysed the sequence of 145 windows running through the sample, each with a fixed size of 120 days and the window itself moving five days at a time. The analysis of the log-returns is chosen because there is a rather weak dependence on the window size compared to the size effects when using prices. It should be noted that the original motivation has been to map the almost continuous space of the returns to the space of discrete variables $(i^*, lag^*)_t$. This may be mathematically viewed as a symbolic map of the time series (see e.g. Crutchfield and Young, 1989, p. 105), which has been an active research field in the past decades. The results presented in Figure 3 indicate relatively lengthy periods with tendencies to prefer $lag^* = 1$, with rare transitions to the region limited by the upper and lower bounds.

7 Conclusion

In this paper, we study the existence of the lead-lag relationships in the dynamics of open ended funds and an all-share index as the representative of capital market. The results stand in stark contradiction to the weak form of market efficiency. Several innovative approaches have been combined with more classic ones, with the study leading to many opportunities for further research. We plan to further focus on the ownership structure of funds and confront it with the links created by the statistical threshold criteria for the selection of the relevant inter-asset cross-covariances. Of interest will be the overlap of these statistical links with the actual ownership relations. The matching between funds and corresponding portfolios composed from the owned assets can be investigated as well. The practical implications of our work may be seen in a framework of the methodologies of portfolio selection or mutual fund ratings.

References

Basher, S. A. and Sadorsky, P. (2006). Day of the week effects in emerging stock markets. *Applied Economics Letters*, 13(10):621–628.

Chiao, C., Hung, K., and Lee, C. F. (2004). The price adjustment and lead-lag relations between stock returns: microstructure evidence from the Taiwan stock market. *Journal of Empirical Finance*, 11(5):709–731.

Chordia, T. and Swaminathan, B. (2000). Trading Volume and Cross-Autocorrelations in Stock Returns. *The Journal of Finance*, 55(2):913–935.

Conover, C. M. and Peterson, D. R. (1999). The Lead-Lag Relationship between the Option and Stock Markets prior to Substantial Earnings Surprises and the Effect of Securities Regulation. *Journal of Financial and Strategic Decisions*, 12(1).

Crutchfield, J. P. and Young, K. (1989). Inferring statistical complexity. *Physical Review Letters*, 63(2):105–108.

De Medeiros, O. R., De Oliveira, G. R., and Van Doornik, B. F. (2009). Testing for Lead-Lag Effects Between the American and the Brazilian Stock Markets. *Available at SSRN 1365322*.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2):383–417.

French, K. R. (1980). Stock returns and the weekend effect. *Journal of Financial Economics*, 8(1):55–69.

Hou, K. (2007). Industry Information Diffusion and the Lead-Lag Effect in Stock Returns. *Review of Financial Studies*, 20(4):1113–1138.

Keim, D. B. (1983). Size related anomalies and stock return seasonality: Further empirical evidence. *Journal of Finanical Economics*, 12(1):13–32.

Lakonishok, J. and Smidt, S. (1988). Are Seasonal Anomalies Real? A Ninety-Year Perspective. *Review of Financial Studies*, 1(4):403–425.

Lo, A. W. and MacKinlay, A. C. (1990). When are Contrarian Profits Due to Stock Market Overreaction? *Review of Financial Studies*, 3(2):175–205.

Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives*, pages 59–82.

Theobald, M. and Yallup, P. (1998). Measuring cash-futures temporal effects in the UK using partial adjustment factors. *Journal of Banking & Finance*, 22(2):221–243.

Toda, H. Y. and Phillips, P. C. (1993). Vector Autoregressions and Causality. *Econometrica: Journal of the Econometric Society*, pages 1367–1393.

Toda, H. Y. and Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics*, 66(1):225–250.
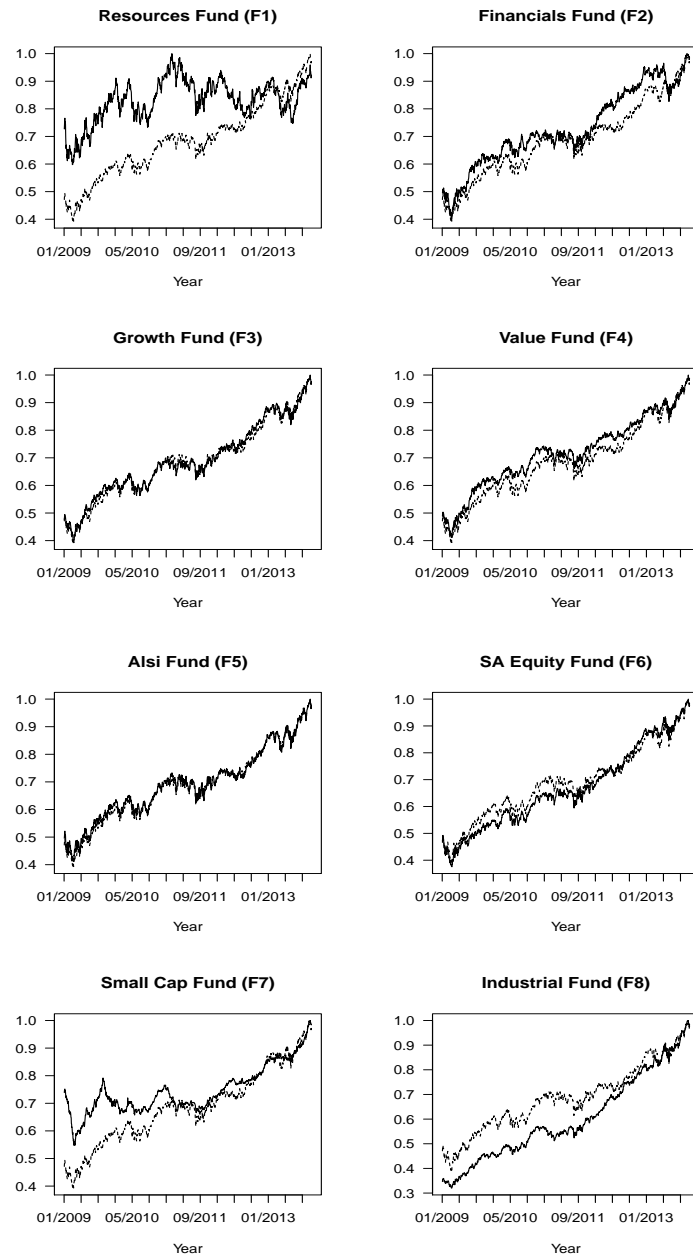
Figure 1: The comparison of the price movement $P_{i,t}^{\mathrm{F}}$ (shown as full lines) of Funds with the dynamics of Index P_t^{I} (dashed lines). The prices are normalized by dividing by the maximal values attained on the entire interval of record. The mentioned normalisation is applied for the graphical purposes only and does not calculations in any way. It can be seen that similar development of prices as Index have Growth and Alsi Fund. It should be noted that Index serves as a benchmark for Growth and Value Fund.
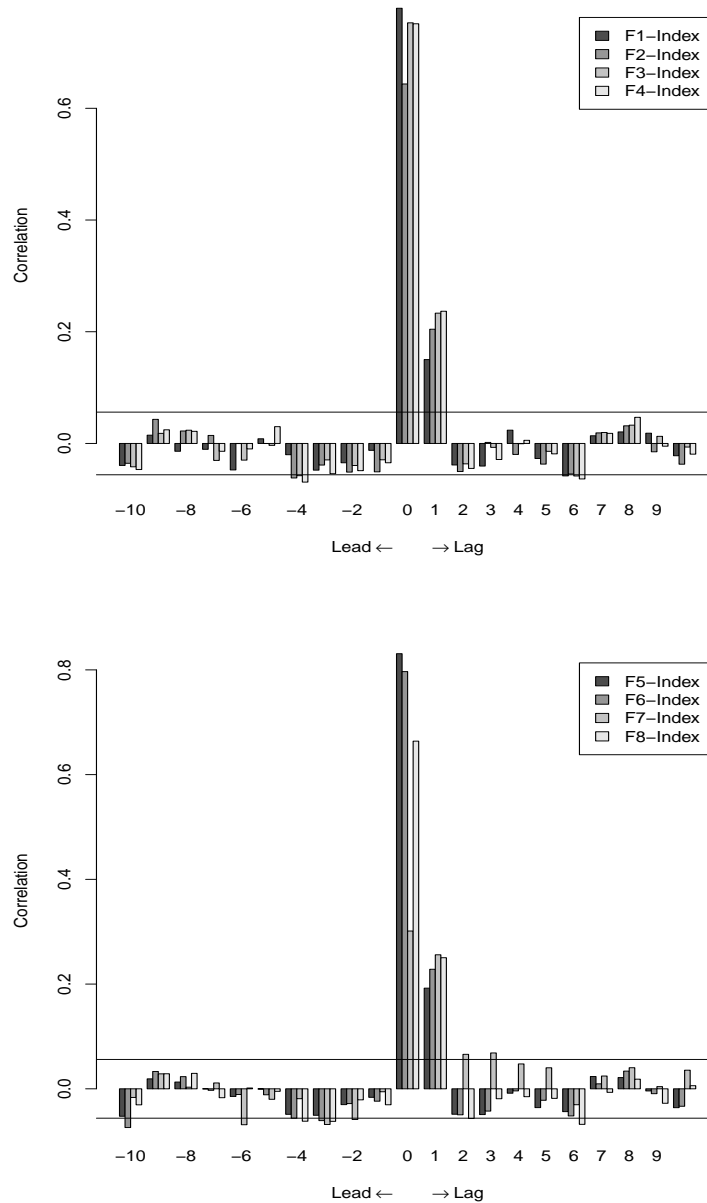
Figure 2: Fund versus Index returns cross correlations calculated for multiple time lags. We see that as expected, the dominating lag relationship corresponds to $lag = 1$, which confirms an existing lead-lag relationship. Lead/lag arrows depict direction of led/lagged returns of Funds to the log-price changes of Index when calculating cross-correlation.
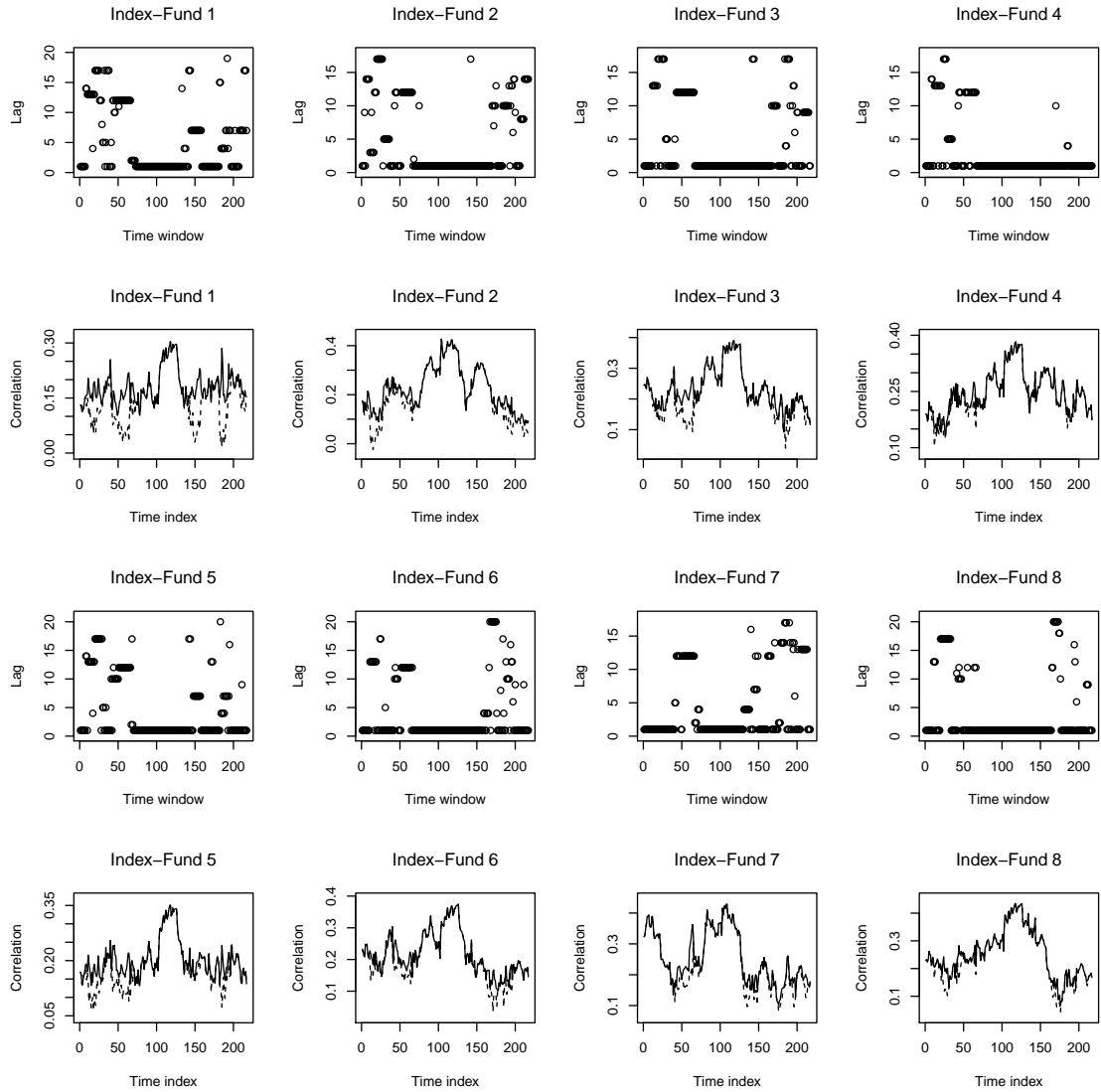
Figure 3: Correlation between log-price changes of Index and returns of Funds when optimal lag^* is used (full lines) and when $lag = 1$ is used (dashed lines). Optimal lag^* is depicted by circles.

12