Reproducing the "Foundations of Technical Analysis": Computational Algorithms, Statistical Inference, and Empirical Implementation" by Andrew W. Lo, Harry Mamaysky and Jiang Wang.

ABSTRACT

In 2000 Lo, Mamaysky and Wang have published a paper in which they show that the indicators of technical analysis(TA) are informative, i.e. that daily stock returns, conditioned on TA indicators are different from daily unconditional stock returns.

The paper is readily readable and the authors honestly report about the shortcomings of the statistical tools they engage.

Lo et al use historical market data from 1962 to 1996. So the natural question is whether the TA indicators stay informative during the "Internet era", i.e. when both individual and institutional traders got enough computational and telecommunicational power to exploit market imperfections.

I try to reproduce the results of Lo et al for the timespan from 1995 to 2010 (historical data for DJ30, SP100, NASDAQ100) and for the timespan from 2003 to 2010 (historical data for 10 stocks from DAX). I come to the conclusion that the results are not anymore reproducible. Additionally, I show that the shortcomings of statistical tools applied by Lo et al should not be disregarded.

Vasily <<YetAnotherQuant>> Nekrasov

January, 08 2011

finanzmaster ἀt gmx DOT net

http://www.yetanotherquant.de

Method

Following Lo et al I engage kernel smoothing to extract the regularities from the stock price chart. Lo et al use the Gaussian kernel and set the bandwidth parameter to $0.3 \times h$ where h minimizes the cross-validation function. The factor 0.3 is chosen by eye (of several professional technical analysts) in order to avoid oversmoothing.

I, in turn, rely on *np* (Nonparametric kernel smoothing methods for mixed data types) package for R. R is a powerful opensource statistical software, freely available from http://www.r-project.org
The routine *npreg* of the package *np* automatically chooses an optimal bandwidth parameter.

In order to control the regression smoothing quality I let my R-script generate the graphical output at every step.

Looking at thumbnail gallery I can readily identify the abnormalities, at least the gravest.

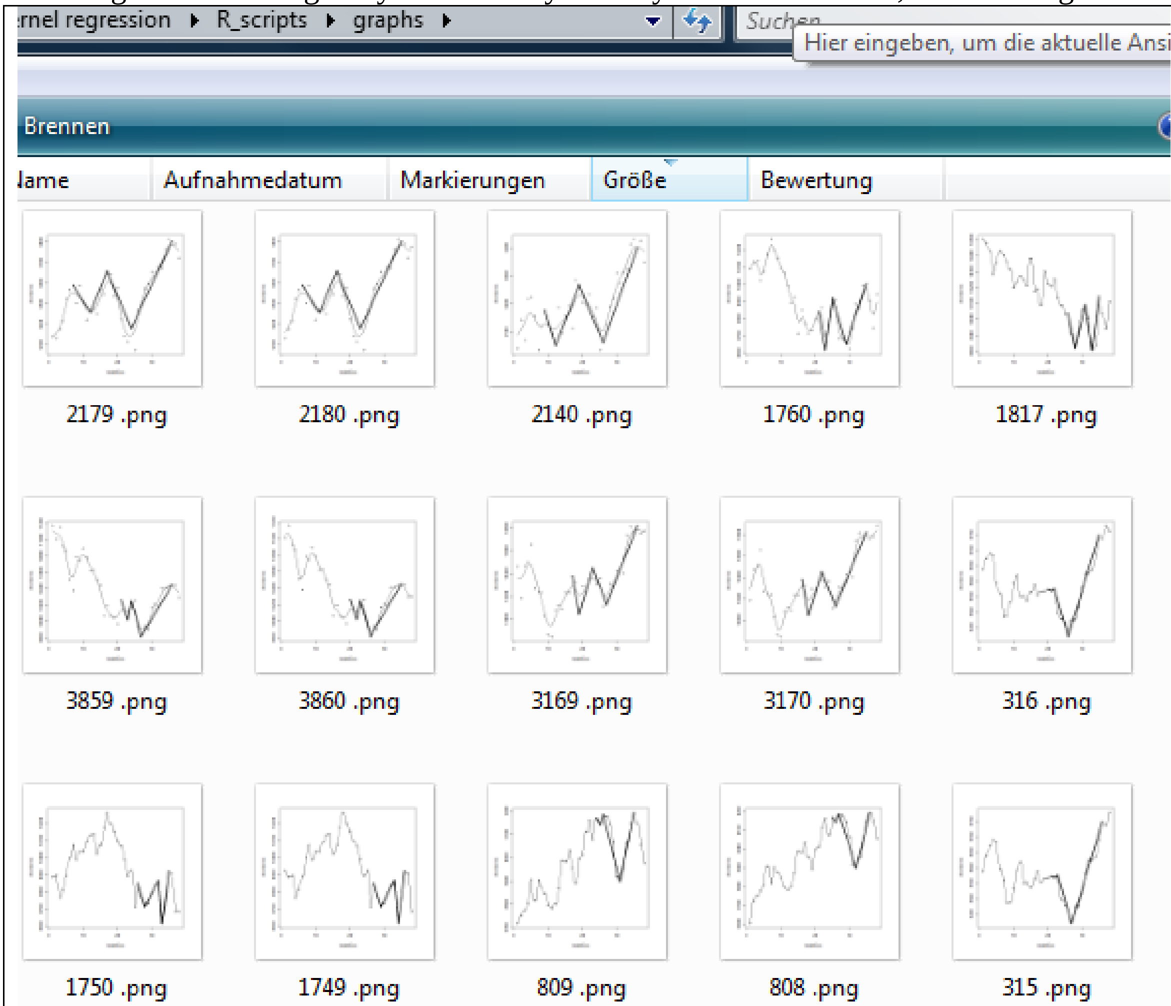


Figure 1: Principle of the quick visual control of results (in this case for BTOP pattern)

In most cases I find the smoothing adequate. Interestingly, when I engage *ksmooth {stats}* (another R routine for kernel smoothing) with the <u>same</u> bandwidth and kernel as in *npreg* I get significantly different results. This lies, most likely, on fine differences in program implementation. In either case *npreg* produces smoother results and much more reliable: *ksmooth* frequently generated no graphical output at all (Figure 3, Step 1017) whereas *npreg* nearly always delivers an adequate picture.

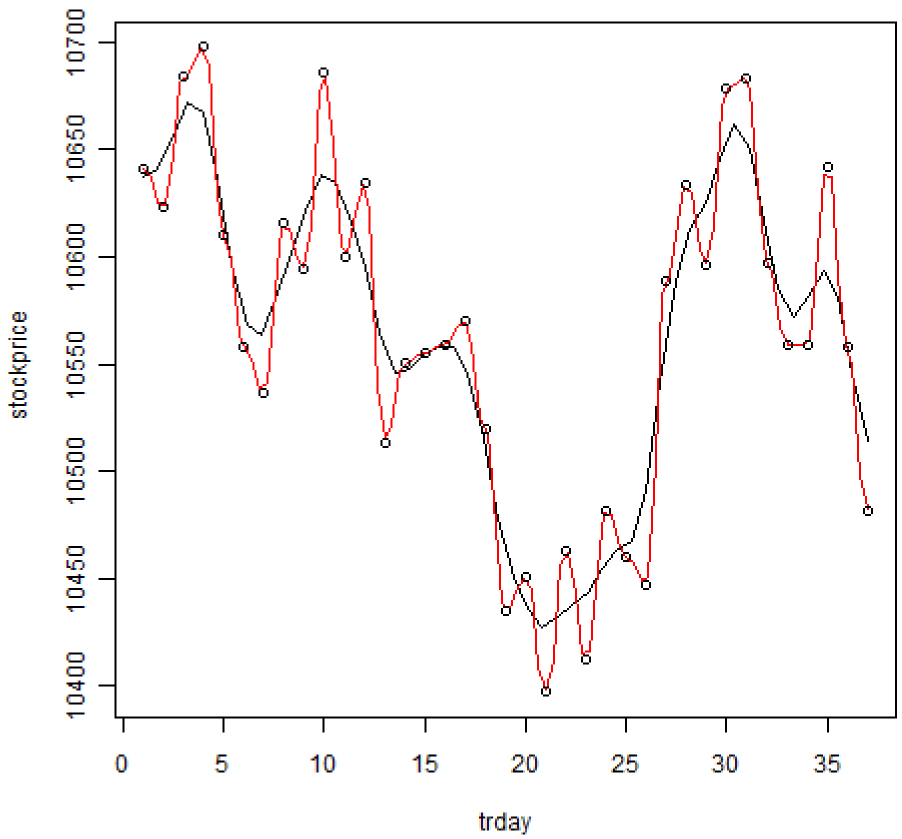
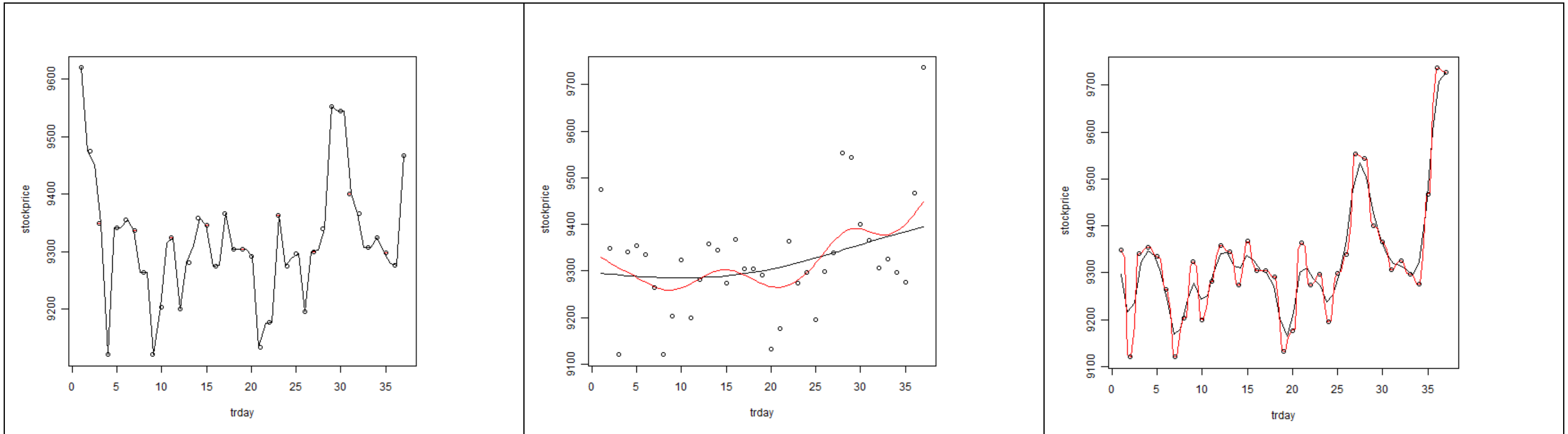


Figure 2: Example of smoothing with *npreg* (black curve) and *ksmooth* (red curve) (DJ30 historical prices, step 2664)

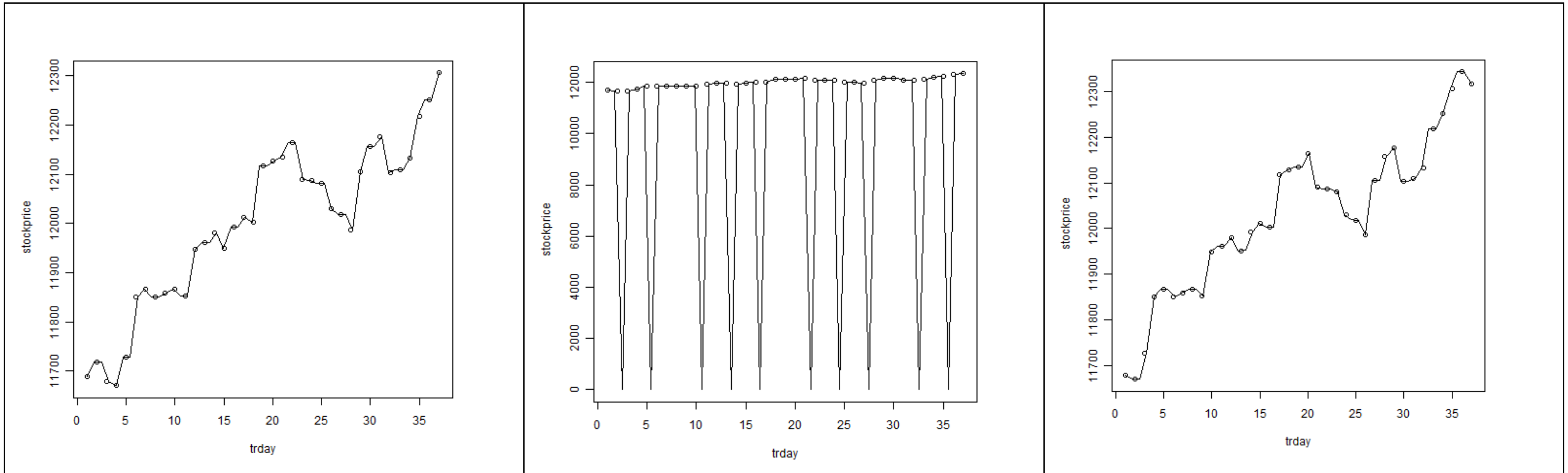
Some abnormalities, however, do appear, namely at step *x* everything is ok, at step *x+1* sudden oversmoothing occurs and at step *x+2* everything is ok again (Figure 3).
It takes place for both *npreg* and *ksmooth* routines.

Trying to multiply the optimal bandwidth with 0.3, as Lo et al do, does not help out. It leads to undersmoothing (Figure 4) which occurs even a little bit more often than an undersmoothing in case of the optimal bandwidth.
For a smaller bandwidth one should expect more pattern recognized. It does take place but the difference in cases of optimal bandwidth *h* and *0.3h* is not so drastic (cp. Table 1 with Table 4). Additionally, with the bandwidth equal to *0.3h* it tends to detecting of too local extrema. That's why I further work with optimal bandwidth *h*.



Step 1017	Step 1018	Step 1019

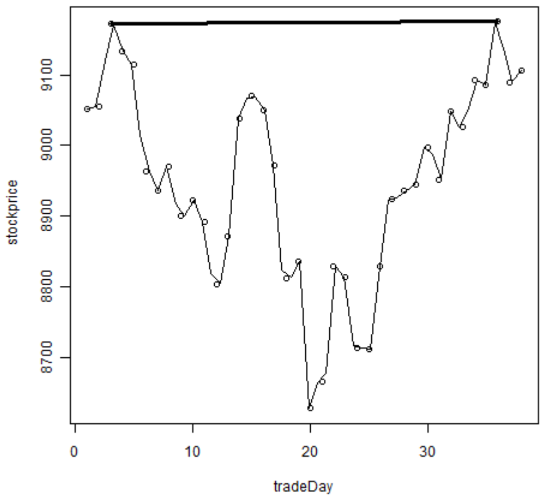
Figure 3. Abnormalities by kernel smoothing. *npreg* (black) and *ksmooth* (red) with optimal bandwidth (DJ30 historical prices)



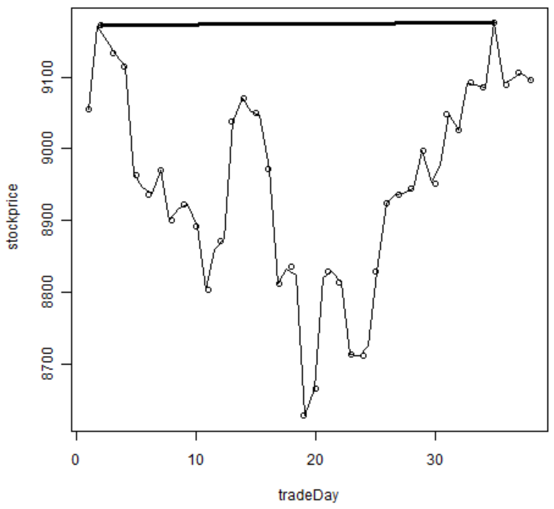
Step 2957	Step 2958	Step 2959

Figure 4. Abnormalities by kernel smoothing. *npreg* with 0.3 x optimal bandwidth (DJ30 historical prices)

Following Lo et al, I set the window length *l*=35 days and the lag *d*=3 which allows for identifying of the last local extremum (i.e pattern completion). Interestingly, almost always it suffices *2* days lag to identify the pattern but I still leave *d*=3. If I get a pattern identified at step *x* and then the *same* pattern at step *x+1* I neglect the latter (Figure 5).



Step 853, DT pattern identified 2 days after completion.



Step 854, the same DT pattern identified 3 days after completion; will not be counted

Figure 5. batch_kernel_regression_DTOP_and_DBOT.r (DJ30.xls)

As to pattern definition, I follow Lo et al with an exception for DTOP and DBOT which I extend as follows:

$$DTOP \equiv \begin{cases} E_x \ is\ a\ \max imum\ and\ occurs\ earliar\ as\ E_a \\ E_x\ and\ E_a\ are\ within\ 1.5\%\ of\ their\ average \\ dis\tan ce\ between\ E_x\ and\ E_a\ is\ 22\ days\ or\ more \end{cases}$$

$$DBOT \equiv \begin{cases} E_y \ is\ a\ \min imum\ and\ occurs\ earliar\ as\ E_b \\ E_y\ and\ E_b\ are\ within\ 1.5\%\ of\ their\ average \\ dis\tan ce\ between\ E_y\ and\ E_b\ is\ 22\ days\ or\ more \end{cases}$$

So the only difference with Lo et all that I use E_x and E_y instead of E_1 hence not missing cases like that on Figure 6. It is worth mentioning that the (counter)example on Figure 6 is artificial, I changed the condition $E_x\ and\ E_a\ are\ within\ 1.5\%\ of\ their\ average$ to $E_x\ and\ E_a\ are\ within\ 0.4\%\ of\ their\ average$ in order to generate it, otherwise the 1st top(bottom) was always the 1st local extremum.

As to search of local extrema, I rely on R package *msProcess*, routine *msExtrema*.

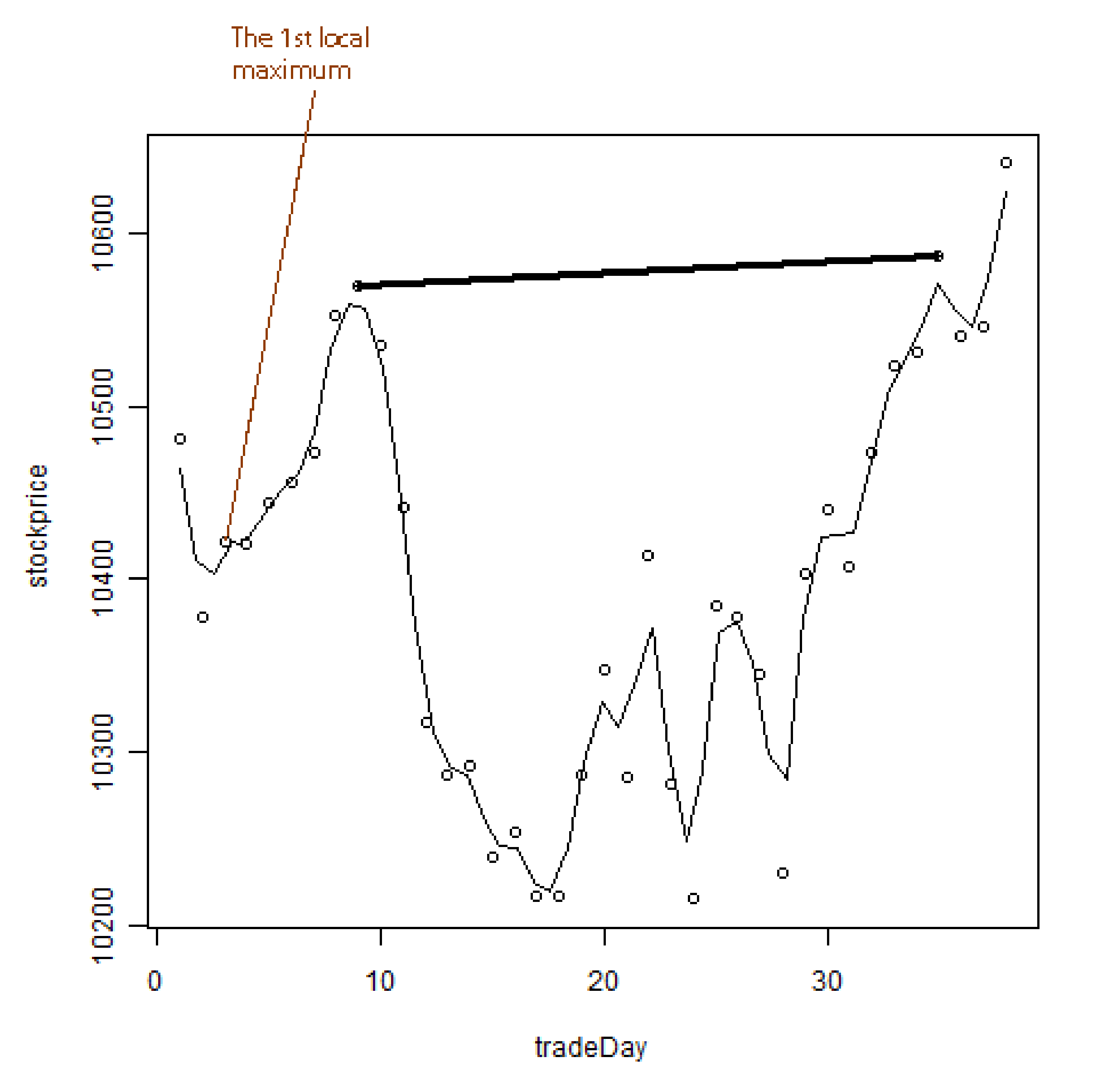


Figure 6. By DTOP the 1st top is not necessarily the 1st local maximum

I do not split the timespan to 5 year periods, partly for simplicity, partly because I believe the timespan 1995-2010 is a "unitary" formation (turbulent time of bubbles and burst). Last but not least, the practitioners trade permanently and cannot do such kind of splitting.

As to goodness-of-fit, I apply two-sample Kolmogorov-Smirnov test to check the null-hypothesis that both unconditional returns and the returns conditioned on patterns come from the same distribution.

I do not apply Chi-square test believing that " Kolmogorov-Smirnov test is more powerful than chi-square test when sample size is not too great" (Ricci).

Results

Table 1. **DJ30** Jan. 1995 - Dec. 2010. **optimal bandwidth**					
Pattern	Number of occurrence	KS-Distance nonnormalized returns	p-value	KS-Distance normalized returns	p-value
TBOT	29	0.1263	0.7476	0.1871	0.2657
TTOP	20	0.1242	0.9189	0.185	0.5037
BTOP	41	0.1536	0.294	0.0672	0.993
BBOT	29	0.1827	0.2914	0.1075	0.8937
HS	145	0.0726	0.4522	0.0976	0.1388
IHS	147	0.0646	0.5951	0.0574	0.7385
RTOP	134	0.1396	0.01276	0.1245	0.03579
RBOT	117	0.0727	0.5857	0.0703	0.6275
DTOP	77	0.1026	0.4045	0.106	0.3641
DBOT	50	0.1276	0.3970	0.0954	0.7592

Table 2. **NASDAQ100** Jan. 1995 - Dec. 2010. **optimal bandwidth**					
Pattern	Number of occurrence	KS-Distance nonnormalized returns	p-value	KS-Distance normalized returns	p-value
TBOT	35	0.1165	0.7344	0.0905	0.9388
TTOP	21	0.1925	0.4211	0.1878	0.4528
BTOP	41	0.2114	0.04874	0.1313	0.4711
BBOT	35	0.2299	0.05104	0.1845	0.1885
HS	111	0.0821	0.4603	0.0848	0.4185
IHS	101	0.1126	0.1645	0.0923	0.3707
RTOP	65	0.1124	0.3941	0.1217	0.2995
RBOT	65	0.1467	0.1276	0.1414	0.1547
DTOP	36	0.1222	0.6613	0.1897	0.1532
DBOT	35	0.1075	0.8179	0.1251	0.6489

Table 3. **SP500** Jan. 1995 - Dec. 2010. **optimal bandwidth**					
Pattern	Number of occurrence	KS-Distance nonnormalized returns	p-value	KS-Distance normalized returns	p-value
TBOT	31	0.1375	0.6064	0.2021	0.1618
TTOP	21	0.1573	0.6795	0.1096	0.9632
BTOP	49	0.1652	0.1422	0.1033	0.6803
BBOT	40	0.1508	0.3289	0.1586	0.2718
HS	153	0.0675	0.5137	0.0852	0.2346
IHS	150	0.0741	0.4054	0.0456	0.924
RTOP	124	0.0944	0.2344	0.075	0.5086
RBOT	107	0.1048	0.2026	0.0946	0.3083
DTOP	74	0.1001	0.4599	0.1062	0.3855
DBOT	45	0.148	0.2836	0.0937	0.8297

Table 4. DJ30 Jan. 1995 - Dec. 2010. **0.3 x optimal bandwidth**					
Pattern	Number of occurrence	KS-Distance nonnormalized returns	p-value	KS-Distance normalized returns	p-value
TBOT	31	0.1185	0.7809	0.1573	0.4316
TTOP	25	0.1563	0.5789	0.1008	0.9625
BTOP	49	0.1987	0.04375	0.0834	0.8894
BBOT	35	0.1289	0.612	0.1113	0.7837
HS	205	0.0618	0.4452	0.0587	0.5133
IHS	207	0.0739	0.2319	0.0855	0.1124
RTOP	203	0.1054	0.02728	0.0866	0.1105
RBOT	175	0.0554	0.6811	0.0484	0.8271
DTOP	80	0.0989	0.4264	0.0815	0.6749
DBOT	45	0.0684	0.9854	0.1197	0.5464

Discussion

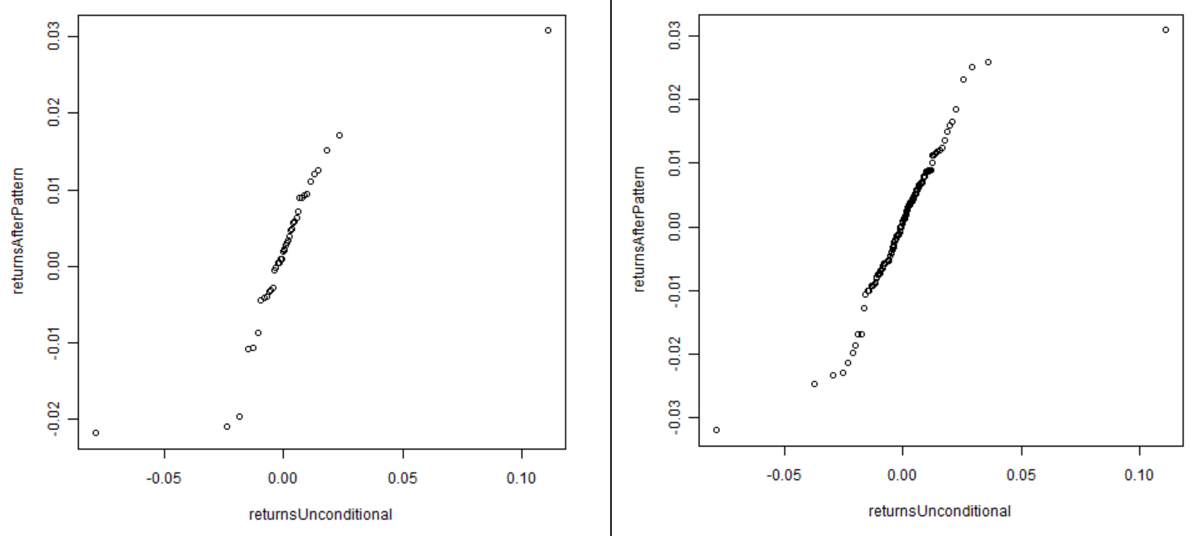
First of all it is worth mentioning that the "power of pattern recognition" of my algorithm implementation is at least not less than that of Lo et al. Let us consider the case of Jow Jones industrial average (**DJ30**).
DJ30 is constituted by the stocks of the companies with the largest capitalization. Lo et al, in turn, consider 50 stocks splitting them to 5 quintiles: from the smallest capitalization to the largest. So it is adequate to confront DJ30 with the largest quintile.
Lo et all consider the timespan from 1962 to 1996 (34 years), I consider the period from Januar 1995 to December 2010 (16 years). So let us compare how many patterns *per year per stock* do we recognize on average. To do this I divide "**My Report . Table 1. Number of occurrence**" by 16 and "**Lo et al**. **Table II. Largest Quintile, 1962 to 1996. Entire**" by 10 stocks * 34 years = 340.
The results are summarized in the following Table 5.

Table 5. Comparison of the recognition power

Pattern	Number of recognized patterns	Power of recognition (per year, per stock)	Number of recognized patterns	Power of recognition (per year, per stock)
	me		Lo et al	
TBOT	29	1.8125	214	0.62941176
TTOP	20	1.25	208	0.61176471
BTOP	41	2.5625	108	0.31764706
BBOT	29	1.8125	110	0.32352941
HS	145	9.0625	208	0.61176471
IHS	147	9.1875	215	0.63235294
RTOP	134	8.375	196	0.57647059
RBOT	117	7.3125	250	0.73529412
DTOP	77	4.8125	308	0.90588235
DBOT	50	3.125	282	0.82941176

Further is clear that in the most cases the p-values, which I obtain, are not small enough to reject the null hypothesis that the unconditional returns and the returns conditioned on patterns come from the same distribution. For the optimal bandwidth the only exception is RTOP and for 0.3*optimal bandwidth are BTOP and RTOP.

Oppositely, Lo et all yield significant KS test statistic for five patterns: HS, BBOT, RTOP, RBOT and DTOP.
Moreover, another [based on quantile comparison] goodness-of-fit test, which Lo et al implement, confirms the significance of all patterns for NASDAQ (**Lo et al.Table VI**) and seven of 10 patterns for NYSE/AMEX (three exceptions are BBOT, TTOP and DBOT).
Such results of quantile-comparison-goodness-of-fit test can be easily explained if one takes a look at QQ-Plots. At Figure 7 there are two examples: BTOP and IHS but QQ-Plots for other patterns look quite similar.



BTOP	IHS

Figure 7. Example of QQ-Plots show discrepancy in tails and good conformance in body of conditional and unconditional distributions

One readily sees it are the tails(extreme events), i.e. the 1st and the last quantiles, what contributes to the discrepancy between the distributions.
I can suggest two explanations: first of all extreme events are relatively rare. Either the patterns are, thus it is not implausible that there no extreme events in the distributions, conditioned on the patterns. Second explanation maybe that extreme events do not occur [directly] after a pattern completes, however, there is no plausible motivation behind this assumption.

As to KS test (which puts more value to the body of the distributions than to their tails) the difference between Lo's et al results and mine can lie on sample size. Combining 10 of 50 stocks in one quintile Lo et al obtain more data, so more power of KS test. However, they normalize returns on individual stock by subtracting means and dividing by standard deviation

$$X_{ij} = \frac{R_{ij} - Mean[R_{ij}]}{SD[R_{ij}]} \qquad \text{(Lo et al, (21))}$$

Tables 1 - 3 show that such normalization substantially influences KS statistic and its p-value. Though the bias direction is not visible with naked eye but in 18 of 30 cases the p-values for normalized distributions are larger, i.e. probability to reject the null hypothesis decreases. It is plausible for distributions that are not too far away from normal, since if we normalize in such a way two normal distributions with different μ and σ, they become identical.

Last but not least, combining the normalized returns of individual stocks into quintiles may increase the power of KS test but since the stock returns are usually correlated, the increase may be not substantial.
Combining the conditional returns on DJ30, NASDAQ100 and SP500 I obtain the following Table 6

Table 6. KS test statistic for combined normalized returns (DJ30 + NASDAQ100 + SP500)		
pattern	**KS-Distance**	**p-value**
TBOT	0.1082	0.2199
TTOP	0.0882	0.7234
BTOP	0.067	0.6007
BBOT	0.0957	0.3010
HS	0.0591	0.1255
IHS	0.0461	0.386
RTOP	0.0786	0.04084
RBOT	0.0805	0.05162
DTOP	0.0929	0.08344
DBOT	0.0589	0.7641

There is no improvement of statistical significance. However, the indices are very strongly correlated, so probably one should try relatively uncorrelated individual stocks.

So as next try I select 10 stocks from DAX (German analogue of Dow Jones industrial average) so that they represent different branches. Respectively, we can expect moderate correlation between these stocks.

Company	**Branch**
Adidas	Commodities (Clothes)
Allianz	Insurance
Daimler	Automobiles
Deutsche Bank	Banking
Deutsche Telekom	Telecommunications
Heidelberger Zement	Building materials
Henkel	Commodities (household goods)
Merck	Pharmacy
RWE	Energy
ThyssenKrupp	Metals

The period I consider is from January 2003 to December 2010 (earlier data are not available from de.finance.yahoo.com)

I do not report the results for each stock since there is nothing special about them and combined normalized returns are as follows in Table 7:

Table 7. KS test statistic for combined normalized returns of 10 stocks from DAX		
pattern	**KS-Distance**	**p-value**
TBOT	0.073	0.1943
TTOP	0.0727	0.2696
BTOP	0.0822	0.1337
BBOT	0.0741	0.3665
HS	0.0573	0.05796
IHS	0.0345	0.5524
RTOP	0.0624	0.1731
RBOT	0.052	0.3593
DTOP	0.0574	0.5841
DBOT	0.1057	0.0671

Nothing special as well.

Finally, I check whether I can obtain the significant statistic for earlier market history, trying DJ30 for a period from 1962 to 1997 (Table 8). There are at least three patterns with significant test statistic, so it might support the assumption that TA-patterns were informative but got exhausted in Internet era.

Table 8. **DJ30** Jan. 1962 - Dec. 1997					
Pattern	Number of occurrence	KS-Distance nonnormalized returns	p-value	KS-Distance normalized returns	p-value
TBOT	74	0.099	0.4682	0.0593	0.9586
TTOP	56	0.1662	0.0924	0.1595	0.1176
BTOP	96	0.0855	0.4912	0.0582	0.9044
BBOT	69	0.092	0.6079	0.0836	0.7245
HS	391	0.0373	0.673	0.0456	0.4177
IHS	396	0.0517	0.2624	0.0525	0.2458
RTOP	338	0.0637	0.1422	0.0406	0.6553
RBOT	319	0.0807	0.03604	0.0454	0.5489
DTOP	180	0.1225	0.01	0.0833	0.1723
DBOT	127	0.1521	0.006084	0.0824	0.3633

References

Lo, Andrew W., Harry Mamaysky, and Jiang Wang, 2000. Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation. The Journal of Finance, 55(4), 1705–1765.

R, np package. Reference manual: http://cran.r-project.org/web/packages/np/np.pdf
Vignette: http://cran.r-project.org/web/packages/np/vignettes/np.pdf

Ricci, Vito Fitting distributions with R: http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf

http://de.finance.yahoo.com