

Looking for short term signals in stock market data

A. Bocharov

Microsoft Corp., Redmond, USA

Abstract

This paper offers a method for generating discrimination criteria that can be used to identify strong patterns in stock market data. More effective criteria are necessary as short-term trades are often driven by noise rather than valid market signals - a viewpoint known as the "noisy trader" model of the stock market. To demonstrate why traditional technical analysis falls short, we applied the standard methods to a synthetic market generated by a random walk. Although the market data is engineered to be pure noise, technical analysis "discovers" strong features such as accumulation/distribution patterns, upward and downward trends, support and resistance levels. These secondary features represent noise in the technical indicators.

By comparing the performance of these indicators between a real market and the synthetic noise market, real signals can be identified. We select a subset of technical indicators $\{T\}$ such that each indicator T in this subset has a mesokurtic (usually, near-normal) distribution across the synthetic noise market. It turns out that many of the common measures (e.g. accumulation, relative strength, moving average trend direction) are mesokurtic for the noise market, but have strongly leptokurtic distributions across the actual stock market. We propose a method for using the kurtosis of such indicators T to separate the noise from the real signal in the indicators. Specifically, having obtained a mesokurtic baseline that corresponds to noise, we use the empirical real market distribution for indicator T to compute discrimination thresholds beyond which a value of T should be considered a strong signal.

The strong signal concept is validated using historical price/volume data for the Standard & Poor's 1500 (S&P 1500) set of stocks.

Keywords: market analysis, kurtosis risk, random walk, technical analysis.



1 Introduction

This paper proposes conceptually simple methods for assessing the predictive power of certain stock market technical indicators and for locating ranges of values in which the indicator values are statistically significant. To accomplish this, we apply statistical analysis to answer the following question: what is the relative likelihood that the observed value of a technical indicator could not be explained by a noisy random walk market model? The analysis is based on the empirical observation that certain indicators are strongly leptokurtic across the actual market and generate numerous outlying values that are highly unlikely under random market hypothesis.

The idea of using leptokurtic distributions for describing capital markets' statistics is not entirely new. Bollerslev and Hodrick [1] explored the fat tailed distributions of stock price returns in conjunction with the market efficiency hypothesis. (See also discussions in [2] and [3]). Fischer [4] proposed the use of generalized secant hyperbolic distribution to describe asset returns. Underestimating the kurtosis of the actual asset distributions poses multiple risks, whereby the term "kurtosis risk" has been coined. The notion has been summarized by Mandelbrot and Hudson in [5]. Recently Kitt and Kalda published an outline of Leptokurtic portfolio theory [6] that relies of returns' kurtosis for minimization of portfolio drawdown risk.

This research takes the above ideas in a different perspective. First, we apply the analysis across the market breadth (for a market of Standard & Poor's 1500 [9] size) instead along the price/volume time series. Second, it is extended to a number of technical indicators. We posit the normality (it is an easy pun to say that the real market is "abnormal", but the irony of it is that its "normal" part maybe indistinguishable from noise) of the random market using the following time/breadth equivalence principle: an indicator that is normally distributed along a random walk should be also normally distributed across the broad market snapshot. The intuition for this principle is that the same kind of interaction of "noisy" market agents that caused a stock to "random walk" would cause similar kind of a random scatter of price and volume across the market. The assumptions regarding the behaviour of indicators in random market are detailed below in the "Technical indicators used" section.

Let T be a technical indicator and let $ND[T,d]$ be the best normal distribution fit for the empirical distribution of indicator T across the actual market on a given date d . On most days covered in this research there will be a significant number of values of T that are highly improbable given $ND[T,d]$ (for example values outside the six-sigma confidence interval). In general, if α is a chosen confidence level and $I_\alpha = [u,v]$ is the respective confidence interval for $ND[T,d]$ then u and v are regarded as strong signal thresholds. A value of T that is either smaller than u or larger than v is a *strong signal in T* at confidence α .

There were six technical indicators employed in this research. We have seen 160 or more values of these indicators across the S&P 1500 on average per day that were outside the respective 99% confidence intervals. A random sampling of



these values was cross-checked against the available news, press releases and earnings data (by human review process which was not automated at this stage). We have found that in most cases the outlying values of indicators appeared to be responses to recent events.

Often strong signals of the day clustered together to highlight movement of money between sectors, as well as sector upgrades/downgrades.

We have also observed the *temporal clustering* of signals, whereby strong signals in an S&P 1500 issue XYZ would be seen for several consecutive days.

An example of a day worth of strong signal analysis for 30th November 2007 is given in the “Case Study” section below.

The main result of this paper is an empirical proof of concept based on statistical analysis of historical price/volume for S&P 1500. This Author is satisfied that measuring the strength of technical indicator values based on the excess kurtosis of the empirical distributions for indicators is a viable method for detecting strong signal issues in the stock market. This is only the first step towards in-depth market signal analysis (as outlined in the “Further Directions” section).

2 Technical indicators used

2.1 Moving average trend strength: relative and logarithmic

To measure the trend strength in the underlying use k-day moving averages of the closing price, denoted [k]dMA. The strength measure that is most amenable to the statistical analysis is

$$TS[k](date) = [k]dMA(date) / [k]dMA(previous\ trading\ date)$$

$TS[k](d) > 1$ iff k-day moving average is trending up on date d and $TS[k](d) < 1$ iff it is trending down. Under the assumption of the “random walk” market $TS[k]$ is log-normally distributed and thus $\log(TS[k])$ has normal distribution.

The sufficient statistic for the short-term signal research here has been collected for $\log(TS[5])$ and $\log(TS[20])$ and it turned out that either of these measures is leptokurtic across the actual market on most trading days.

2.2 Relative strength: log(RS)

Relative strength (RS) is the basis of the widely used Relative strength index (RSI, [7]). For a chosen moving window, the relative strength of a stock price curve is defined as $RS = (\text{Average gain})/(\text{Average loss})$.

For our purposes we used a 20 day accumulation window (which was found empirically to be more robust than the usual 14 day window).

For a “random walk” market RS is distributed log-normally, so it is **Log(RS)** that was the indicator most suitable here. The **Log(RS)** appears to be leptokurtic across the real market with typical excess kurtosis between 0.3 and 1.7

2.3 Accumulation rate (normalized)

For a chosen accumulation window of n OHLC bars, the accumulation/distribution rate is defined as



$$AR = \sum_{i=1}^n \frac{(2Close_i - Low_i - High_i)}{(High_i - Low_i)} Volume_i$$

(cf. [8]). The underlying is said to be “under accumulation” if AR is positive and to be “under distribution” if it is negative.

While this measure makes perfect sense along a single price/volume time series, for cross-market comparison the rate must be normalized to trading volume, so that comparison between high volume and low volume securities is possible. Therefore the measure actually used was $ARN = \frac{AR}{V_a}$, where

V_a is the mean trading volume.

Strictly speaking ARN is not normally distributed over a “random walk” market. The core model for this indicator there is $\sum_{i=1}^n U_i V_i$ where all the variables are independent, each U_i is uniformly distributed on the [-1,1] and all V_i have the same log-normal distribution. However, for a reasonable choice of the random market assumptions, this distribution is going to be mesokurtic, typical excessive kurtosis not to exceed 0.1

2.4 “Day trader index” and ROI

A new indicator introduced specifically for this research is parameterized by a “trading window” of n days and a “trading threshold” t, which is a small positive number, typically between 0.005 and 0.015. For a daily OHLC bar of a given price curve on i-th day define the

$$P_i = \frac{(High_i - Open_i)}{Open_i}; L_i = \frac{(Close_i - Open_i)}{Open_i}$$

and

$$DR_i = \text{if}(P_i > t) \text{ then } P_i \text{ else } L_i$$

Now define $DTI = \sum_{i=1}^n DR_i$

DR_i rates a day as “losing” if the price never rises high enough above the opening price and the closing price is lower than the opening one. The day is rated “winning” otherwise. Therefore DR_i scores a win for a day when the underlying had sufficient upward momentum. Conceptually DTI represents returns for a severely idealized model day trader, who always buys at Open and never holds past Close. Technically the indicator relates to both the upward momentum of the underlying and its benign intraday volatility.

Along with the DTI this Author also used a more sophisticated index generically dubbed ROI which does not have a closed form description and had



been computed dynamically. ROI related to a trader model that allowed overnight holding and used adaptive day-to-day adjustments of the threshold t . It is somewhat amusing that on average across S&P 1500 DTI and ROI were tied.

3 Signal strength and signal clustering

If $N(\mu, \sigma)$ is the best normal distribution fit for empirical distribution of indicator T across the market then we define the *strength of signal* $T=v$ as

$$S(T = v) = \left| 2\Phi\left(\frac{v - \mu}{\sigma}\right) - 1 \right| \text{ where } \Phi \text{ is the probit function.}$$

As per this definition the strength of values close to “market mean” $T = \mu$ is close to 0, whereas the strength of a value on either fat tail of the empirical histogram would be close to 1. For example, $S(T = \mu + 3\sigma) = S(T = \mu - 3\sigma) = 0.9973$.

An important concept that requires further research is the *joint signal strength* for a collection of signals. Generalizing the above, the joint signal strength should be derived from the joint cumulative normal proxy of the signals. The complication here is that all indicators are computed from the same body of price/volume data and therefore cannot be automatically considered independent.

For a selected strength threshold t we may see strong signal(s) in one or more indicator for a stock XYZ for L consecutive trading days starting date d . When this is the case, we say that the stock exhibits *strong signal clustering sequence* of length L . For $L > 1$ it is usually the case that the number and strength of signals increases then decreases as the clustering sequence unfolds.

4 Results: indicator and clustering statistics

The signal strength analysis has been run on 12 months worth of historical S&P 1500 day OHLC/Volume price/volume quotes bars.

We have found that the logarithmic measures based on Relative Strength, Trend Strength for 5-day and 20-day Moving Averages are strongly discriminating in that they practically always leptokurtic across the actual market. The indicators (and/or their combinations) can be therefore used to measure the strength of signals on most trading days. DTI and ROI were leptokurtic on about 92% of days.

Table 1 describes typical monthly changes in excess kurtosis of indicators used in this research. As one can see the normalized A/D is not strongly leptokurtic, and thus should be used as an auxiliary measure on days when its kurtosis is sufficiently high.

A strong signal in a particular stock would be a part of *temporal signal cluster* in approximately 50% of all instances. The durations of such signal clusters have a highly skewed distribution that tails-off exponentially. An example of a histogram for strong signal lifetimes in a month window is shown in Figure 1.



There were overall 660 incidences of strong signals at confidence level 99% in December 2007. Among these, 343 instances lasted for just one day. On the right end of the spectrum there was one stock (IMB)(IndyMac Bancorp’s year-long meltdown has been sealed on Dec 11 when its credit rating has been cut to Junk) that exhibited strong signals for 15 consecutive days.

Table 1: Monthly changes in excess kurtosis of indicators used.

Indicator	Log(RS)	Log(TS[5])	Log(TS[20])	Norm. A/D	DTI	ROI
Mean Ex.Kurtosis	1.014	9.43	6.0	-0.075	9.1	0.9
Ex. Kurtosis StDev	0.345	3.46	1.93	0.28	6.4	0.64

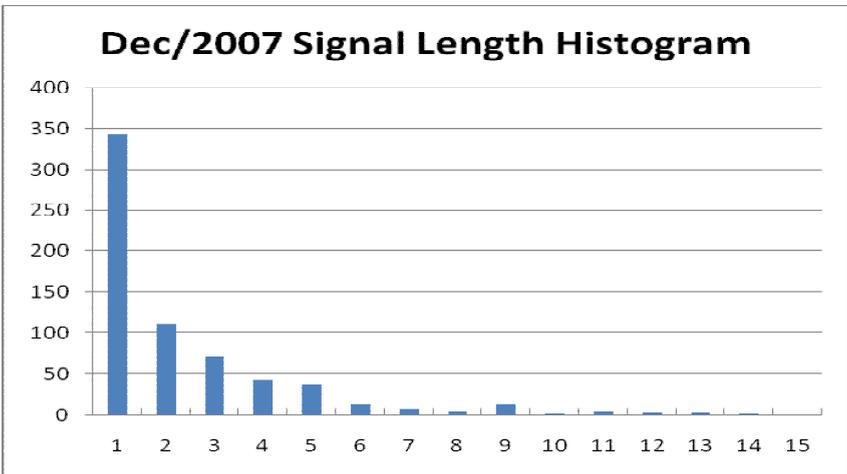


Figure 1: Strong signal lifetimes in December 2007 (length in days).

5 Case study

For case study we were looking for outlier values of indicators that were outside the 99% confidence intervals of the respective empirical normal distributions, i.e. signals with strength 0.99 or higher. There were around 160 outlying values registered per day for the S&P1500 data (on average), so the study was limited to symbols with more than one strong indicator value.

Example . November 30 2007

The 11/30/2007 has been randomly selected for detailed discussion. The market on that day was unremarkable in terms of direction, yet offered a broad variety of signals of many interesting types. Table 2 below highlights 35 issues with multiple strong signals. (There were more than 90 issues, showing strong signal in just a single indicator, which could not be covered here for obvious lack of space). For 31 of the issues in table 2 the reasons for the strong signals are



Table 2: Strong market signals on 11/30/2007.

Symbol	Outliers	Strength	Sign	Apparent reasons *)	Start	End date
AAPL	TS[20],TS[5]	0.99, 0.99	++	iPhone to China, France	11/12	11/30/07
ACAT	TS[5],ROI	0.992, 1	--	“global warming”	11/26	12/06/07
ADS	ROI, A/D	0.991, 0.998	-+	Buyout rumour	11/28	12/04/07
AMD	DTI,ROI	0.998, 0.993	--	Outlook, downgrade	10/18	Unknown
ARM	TS[20],RS	0.994, 0.992	--	Loss, weak USA sales	11/14	12/06/07
ASHW	TS[5],DTI	1,1	++	Trend Reversal (**)	11/19	11/30/07
CBG	TS[5],DTI	1,0.998	++	Shares buyback	11/23	12/03/07
CBM	TS[20],RS	1,0.999	--	Downgrade	11/05	12/06/07
CENT A	TS[5],TS[20], DTI	1,1,1	--+	Trend reversal (**)	11/28	12/10/07
DITC	TS[20],A/D	1,0.997	--	Neg. earnings outlook	11/1	Unknown
DSL	TS[5],DTI	0.999,0.999	++	Investor stake revealed	11/19	12/03/07
ETFC	TS[20],ROI	1, 0.993	--	Subprime exposure	11/12	Unknown
FBTX	TS[20],TS[5]	1,1	--	Neg. analyst coverage, loss outlook	11/19	12/06/07
FINL	TS[5],DTI	0.994, 1	++	Speculation around lawsuit	11/26	12/10/07
FLE	TS[20], TS[5]	1, 0.995	--	Guidance, loss anticip.	11/1	12/07/07
FNM	TS[20], TS[5], DTI	1, 0.998, 0.998	-++	Gov. probe, losses, benign insider buying	10/06	12/06/07
FRE	TS[20], TS[5],DTI	0.999, 1,1	-++	Gov. probe, \$2B losses, good capital plan	10/06	12/06/07
GCO	TS[20],DTI	0.996,1	-+	Lawsuit, insider buying	11/19	12/28/07
GVA	TS[5],DTI	0.99, 0.996	++	Business expansion	11/24	11/30/07
GVHR	TS[20],TS[5], RS,ROI	1, 0.999, 1, 0.999	----	Weak earnings, outlook	11/05	12/03/07
HANS	TS[20],RS	1,0.992	--	Negative earnings	11/08	12/03/07

Table 2: Continued.

ISRG	TS[20],TS[5]	0.995, 0.995	++	Pos. analyst coverage	11/27	12/07/07
LYV	TS[20],ROI	0.996, 0.998	--	Negative earnings	11/08	12/06/07
MDRX	TS[20],RS	1,1	--	Neg. earnings surprise	11/09	Unknown
MHO	TS[20], TS[5]	1,1	--	Housing slump	10/19	12/06/07
MTG	TS[20],TS[5], DTI volume	1,1,1	+++	Change of mortgage standards	11/28	12/10/07
NLS	TS[5],DTI	0.998, 0.994	++	Trend reversal (**)	11/21	12/07/07
PENX	A/D,RS,DTI, ROI, volume	0.993, 0.998, 1, 0.995	----	Plans stock offering	11/28	12/03/07
PMTI	TS[20],RS	0.992, 1	--	Weak earnings	10/25	12/07/07
RAI	RS, A/D	0.99, 0.995	++	PR, Dividend	11/27	12/05/07
SAM	TS[20],RS	1,0.991	--	Negative earnings	11/06	12/05/07
SMRT	TS[5],ROI	0.994, 0.995	--	Downgrade, loss	11/22	Unknown
THC	TS[5],TS[20], DTI	1,1,1	+++	Business deal	11/27	12/10/07
URI	TS[20],RS	0.998, 0.994	--	Lawsuit	11/14	1/10/08?
VTIV	TS[20],RS	0.996, 0.997	--	Falls against positive earnings	11/07	12/06/07 ?

*) Apparent reasons: recent event(s) bound to cause the price/volume change.

***) Rebound from apparent support level in absence of major published news.

clear from the historical data. Trend reversals in ASHW, CENTA, NLS happen in absence of recent publicly available news, so the reversals, while technically valid, lack fundamental explanation as of this writing. VTIV is the most interesting of the four ambiguous rows. Despite positive earnings surprise on 11/7/2007 VTIV shares slid 25% over the next few days and stayed depressed until at least 12/17/2007. There were no public disclosures or sector events to explain the decline. The issue does not appear to be fully valued. This may point to a body of tightly kept investor knowledge defining the price. Overall in 11/30 example the ambiguity is limited to 11.5% of the “strong signal” issues surveyed, and it can be perhaps resolved by using better news sources or looking at subcritical values of other indicators.

Note that many of the patterns identified on 11/30/2007, especially the down trends, break down around the same date of 12/06/2007. This is due to market-wide “sucker rally” of 12/05 – 12/10, 2007 in course of which many issues were



broadly bought, especially the oversold stocks of healthy companies. It is interesting to note also that the November/December strong signals reflect the key developments of the period such as iPhone craze, subprime mortgage crisis, restructuring in finance, rush of smart money into pharma and bio and, amusingly, the global warming.

6 Conclusion

In the context of daily price/volume data, the statistics that we have collected and analysed offers strong evidence that certain (log-transformed) technical indicators have strongly leptokurtic distributions across stock market of S & P 1500 size. When these same indicators are expected to be normally distributed under the time/breadth equivalence principle, the excess kurtosis of the actual empirical distributions can be used to measure the strength of outlying indicator values. Spot checks performed around randomly selected strong signal instances suggest significant correlation between the signals and recent business events. The scope of strong signals registered for an individual stock spanned across a sequence of consecutive days (“signal clustering”) in approximately half of the cases thus suggesting some degree of signal autocorrelation.

7 Further directions

1) Technical indicators used in this research were derived from Relative Strength, Accumulation/Distribution and Moving Average measures. It is important to investigate how other popular technical indicators can be applied for signal filtering and whether using additional indicators provides significant quality lift to the filtering. The key technical step for this is describing joint distributions of technical indicators under the random market hypothesis.

2) It is practically important to investigate whether the above results scale down to the level of hourly price/volume distributions and thus address the problem of measuring the strength of intraday signals.

3) For efficient validation of results in future it is necessary to develop automated news analysis agents in order to automate the search for correlation between strong signals and market news.

Acknowledgements

The Author wishes to thank Michael Trott for discussing theoretical statistical distributions of some indicators, Chris Meek for comments on the concept of this paper and John Bocharov for extensive editorial help.

References

- [1] T. Bollerslev, R. J. Hodrick, Financial Market Efficiency Tests, NBER Working Papers 4108, National Bureau of Economic Research, Inc., 1992



- [2] J. Knight, S. Satchell, Return Distributions in Finance, Reed, 2001
- [3] C. Brooks, Introductory Econometrics for Finance, Biddels, UK, 2002
- [4] M. Fischer, Skew Generalized Secant Hyperbolic Distribution: Unconditional and Conditional Fit to Asset Returns, Austrian Journal of Statistics, Volume 33, 2004, PP.293 – 304
- [5] B. Mandelbrot, R. Hudson, Richard L., The (mis)Behaviour of Markets : a Fractal View of Risk, Ruin and Reward, London : Profile, 2004,
- [6] R. Kitt, J. Kalda, “Leptokurtic Portfolio Theory”, The European Physical Journal B, Volume 50, 2006, PP. 141 - 145
- [7] J. Welles Wilder, New Concepts in Technical Trading Systems , TR, 1978
- [8] M. Chaikin, unpublished, (see http://en.wikipedia.org/wiki/Accumulation_distribution)
- [9] The constituent list of S&P Composite 1500 can be downloaded from the Standard & Poor’s web portal <http://www.standardandpoors.com> . (ANDW, CHZ, EDO, LYO were excluded from this analysis due to S&P 1500 rotations.)

