Automated Sentiment and Finance

Chris Poulin – Poulin Holdings, LLC

Anders Madsen – Hugin Expert A/S

Jagrata Minardi, Jeffery Mergler, and Kellie Wills – Insightful Corporation

March 14, 2008

General Description

Utilizing the predictive analytics core of Patterns and Predictions™ (Poulin and Madsen) our Semantic Signal is additionally comprised of the patent pending AF algorithm (Poulin). The short description of the AF algorithm is that it is a computational linguistic technique that improves the specificity of correlations of corpus elements and entities.

Meanwhile, over the past 5 years we have proven time and again that this technique is more effective than standard Information Retrieval approaches in text based classification. This report documents a real world case study of this signal embedded into a fully working trading system written in S+.

Signal Basics

Our research from 2003 was to measure the relationship of financial news and market behavior. For this we generated two S&P related signals. The S&P 500 is a stock market index containing the stocks of 500, mainly American, corporations. Our signals are of futures contracts derived from this index.

Our internal codes for these signals are;

'585_05_5_2', is our original signal that runs 585 days of S&P futures prices vs. a customized financial news database. The time period here is May 2, 2005 – Jan 31, 2006. This signal presently performs at a consistent **52.33%** predictive accuracy over time.

'005_5_5_25', is an extended signal that runs against approx 7 years of S&P futures prices vs. a large LDC (Linguistic Data Consortium) NLP data corpus. Note: LDC data sets are THE standard in both computational linguistic and search engine back testing. The time period here is Jan 2, 1997 – Dec 31 2003. This signal presently performs at a consistent **52.55%** predictive accuracy over time.

Trading Basics

Our signals were initially submitted to the Insightful Corporation for validation as a 'double blind' procedure in the Spring of 2007. The next phase of the engagement was to build a complete portfolio trading system around the Poulin_Madsen signals, and this work completed on March 14, 2008. Fundamental to the process was the choice of an asset allocation model.

For asset allocation, we followed a Kelly Criterion model. The Kelly criterion is considered a strong long term strategy, but is acknowledged to produce volatile results. "In probability theory, the Kelly criterion, or Kelly formula, is a formula used to maximize the long-term growth rate of repeated plays of a given gamble that has positive expected value. It was described by J. L. Kelly, Jr, in a 1956 issue of the Bell System Technical Journal." (Source: http://en.wikipedia.org/wiki/Kelly_criterion)

Bottom Line

The initial Signal 585_05_5_2 had a return of -21.50%, vs. a 9% S&P gain for 585 days in part due to volatility. However, the improved Signal 005_5_5_25 had a return of ***+57%,*** *vs. a 9% S&P gain for a nearly 7 year period, even when accounting volatility.*

Trading Rules

Two trading rules have been designed with a level of thoughtfulness appropriate to demonstrate the trade-worthiness of the semantic trade signals.

Various features of a practical trading system, such as stop orders, bid limits, and transaction cost constraints, have not been implemented in this phase of development. Further analysis will explore the sensitivity of the semantic trade signals to different trading rules. Nevertheless, in the case of a base signal with nearly 7 years of daily trade history, we have observed quite fair return over the base signal (S&P).

The trading rules are comprised of parameters, allocation calculations, and transaction schedules, and are described below. In both strategies, cash holdings are valued under the assumption that they earn the risk free-rate.

Parameters:

1. Criterion half-Kelly vs. Kelly.
2. Starting capital.
3. Credit limit; factor applied to available (default=0, for no credit).
4. Risk-free rate for cash holdings.

Calculation of amount to trade at open:

1. Determine whether short-selling is allowed.
2. Apply either the Kelly or the Half-Kelly criterion to determine the target amount to potentially trade.
3. Determine whether we exceed the credit limit; trim the trade amount to the credit limit if so.

Strategy 1: Simple day-trading ("Open-to-close"). In this strategy, at the close of each trading day, share holdings are cashed out, the portfolio is valued, and the transaction for the next trading day's open is calculated. The trade is executed on open of the next trading day.

Strategy 2: ("Open-to-open"). In this strategy, at the close of each trading day, the portfolio is valued and the transaction for the next trading day's open is calculated. The trade is executed on open of the next trading day.

Performance Metrics

The routines in this phase of work produce both daily trade information and portfolio summary information. The metrics in the latter file are described below.

Date	date of trading day
Open	open price on Date
High	high price on Date
Low	low price on Date
Close	close price on Date

Delta	daily close price minus daily open price
Signal	predicted weighted direction of movement (lies between -1 and 1)
Position	signum(signal)
Confidence	not used
Confidence2	either passed from signal data set or calculated as up prob minus down prob
Confidence3	not used
Confidence4	not used
Alloc	fraction of Capital to mark for investment
Cash	cash component of daily portfolio
Capital	dollar value of total portfolio (cash and shares), taken at close, and based on starting capital
Return	total portfolio return, as an annualized percent, relative to close
A.Gain	daily change in Capital
Shares.Held	number of shares held in portfolio
Shares.Trade	number of shares traded at open to arrive at current portfolio
NewInvestAmt	dollars brought to market to trade at open, as calculated and constrained constrained by trading rules
InvestAmt	dollar value of shares held, valued at open

Portfolio Performance

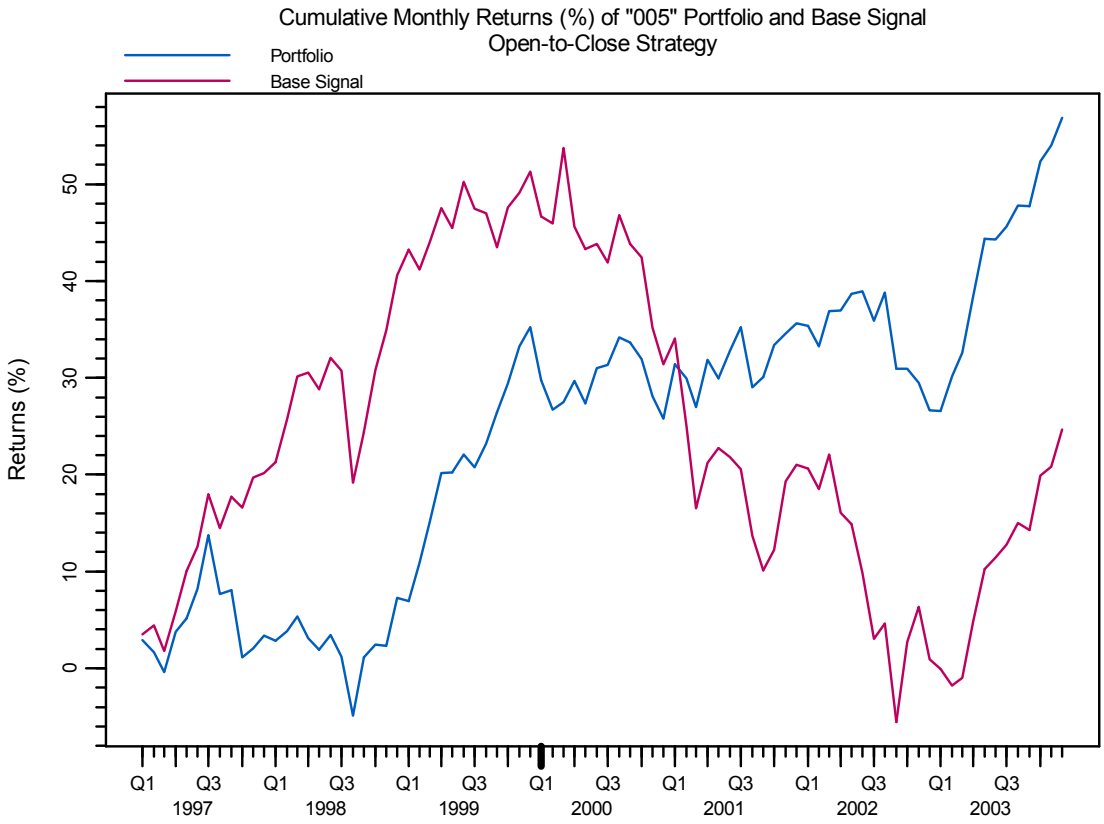
We consider the signals from the files 005_5_5_25.xls and 585_05_5_2.xls, and summarize the result of trading on the signals under the two strategies above. The 585_05_5_2 signal has only roughly 10% of the history found in 005_5_5_25 (172 days vs. 1705 days). There is good news in that we find substantial gain in return over the base signal ("benchmark") for the 005_5_5_25 case. The benchmark is the S&P futures price.

The first two tables below summarize portfolio performance of the signal in 005_5_5_25.xls with respect to the two strategies. The benchmark in this case gains about 1.35% annually over about 7 seven years. When daily returns are calculated from open to close only, the benchmark gains about 3.61% annually. Under both strategies, the portfolio outperforms the benchmark, and the excess returns show lower downside risk than either portfolio or benchmark.

Currently, the 585_05_5_2 analysis shows that the portfolio does not beat the benchmark in either strategy. The Sortino ratios for portfolio and benchmark are somewhat smaller than in 005_5_5_25 for both strategies. Note that the relative sizes of the Sortino ratios between portfolio and benchmark shift places between strategies for the 585_05_5_2 signal. The same is true for downside risk. This reflects the nature of off-hours price movements. A careful investigation of this distinction between the strategies would consider both the total returns and return volatility due to after-hours movements.

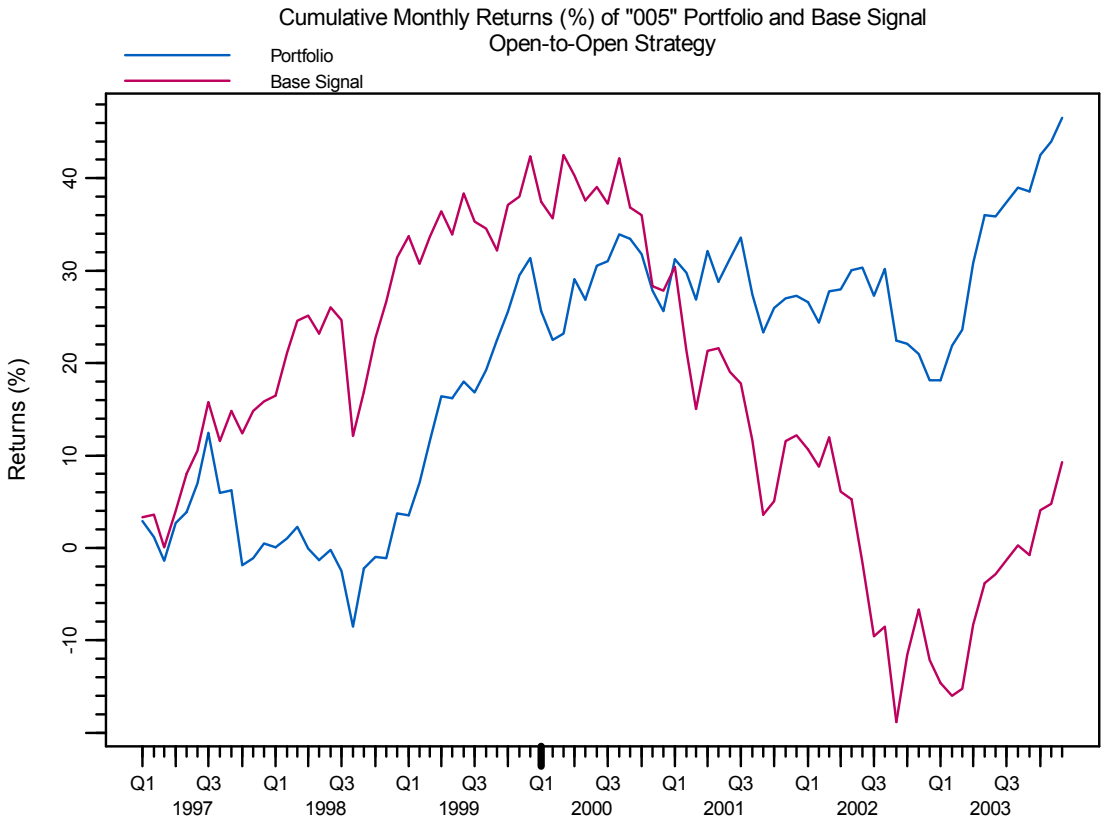
There is a confounding between the effect of the strategy and the behavior of the after-hours price movements. To sort this out going forward, we will consider working in all cases with an instrument, such as SPX futures, that is traded 24 hours a day.

On the following page we will discuss the 005 signal (portfolio) vs. S&P (base signal).



Strategy 1 (open-to-close) on 005_5_5_25:

	Portfolio Return	Benchmark Return	Excess Return
Min	-1476.88	-1476.88	-1299.27
Max	989.02	1302.27	1381.59
Mean	8.31	3.61	4.71
Median	3.00	6.84	0.00
q01	-572.76	-673.88	-604.99
q05	-324.59	-465.83	-302.42
q25	0.48	-149.49	0.00
q75	43.82	162.51	0.66
q95	315.95	435.01	364.38
q99	539.51	721.44	594.98
Var	37604.06	76891.07	38670.01
StDev	193.92	277.29	196.65
MAD	108.21	207.27	103.77
Skewness	-0.42	-0.07	-0.11
Kurtosis	7.25	2.13	6.92
JB.p.value	0.00	0.00	0.00
Sharpe	0.03	0.00	0.01
Omega	1.09	1.00	1.02
mthDnsddv	41.16	68.74	37.64
mthSortino	0.13	0.01	0.05



Strategy 2 (open-to-open) on 005_5_5_25:

	Portfolio Return	Benchmark Return	Excess Return
Min	-1476.88	-1672.09	-1301.29
Max	989.02	1314.06	1675.09
Mean	6.80	1.36	5.45
Median	3.00	8.01	0.00
q01	-580.70	-741.96	-699.39
q05	-328.09	-473.67	-334.67
q25	0.54	-149.30	-12.08
q75	41.45	157.44	19.73
q95	320.87	448.48	373.46
q99	567.51	781.86	616.52
Var	39181.13	84421.72	44500.68
StDev	197.94	290.55	210.95
MAD	108.78	212.18	111.67
Skewness	-0.53	-0.22	0.25
Kurtosis	7.63	3.07	9.08
JB.p.value	0.00	0.00	0.00
Sharpe	0.02	-0.01	0.01
Omega	1.06	0.98	1.03
mthDnsddv	43.22	70.05	36.82
mthSortino	0.09	-0.02	0.07

Headless Implementation and Trade Order Generation

In this phase of development, both the back-testing and daily update routines were parameterized and constructed to run in batch or "unattended" mode, in anticipation of deployment use cases. In connection with this, an XML format for writing trade orders to file was identified. The FIXML protocol (http://www.fixprotocol.org) is used by many in practice today.

Below is an annotated xml stream representing a trade in the FIXML protocol. This can be generated very easily from the current system.

Tag/Attribute	Meaning
<FIXML>	Root element
<NewOrdSingle ClOrdID="123456" Side="2" TransactTm="2001-09-11T09:30:47-05:00" OrdTyp="2" Px="93.25" Acct="26522154">	New order Client's order ID Sell order Transaction time Limit order Limit price Customer's account
<Hdr Snt="2001-09-11T09:30:47-05:00" PossDupFlag="N" PossResend="N" MsgSeqNum="521"> <Sndr ID="AFUNDMGR"/> <Tgt ID="ABROKER"/> </Hdr>	Header element Sending time Possible duplicate flag Resend flag Message sequence number Buyside's CompID Sellside's CompID End of Header Element
<Instrmt Sym="IBM" ID="459200101" IDSrc="1"/>	Stock symbol Stock CUSIP (ID type=CUSIP)
<OrdQty Qty="1000"/>	Order quantity
</NewOrdSingle>	Close of order
</FIXML>	Close root element

Appendix A. Glossary of Terms

To aid discussion of the system both internally and externally, some key terms are defined in this section. In general, we have followed Amenc, Malaise, Martellini, Vaissie; *"Fund of Hedge Fund Reporting: A Return-Based Approach to Fund of Hedge Fund Reporting"*; Edhec Business School, February 2004.

Returns. Continuous (geometric) returns. Unless otherwise noted, returns indicate total returns stated on an annualized basis.

Excess returns. Difference between the returns of an asset and those of a reference benchmark. The reference benchmark is often a proxy for the market, such as the S&P 500 index.

Mean Absolute Deviation ("MAD"). A symmetric measure of variability that is more robust against outlying values than the classical standard deviation.

Jarque-Bera test. A statistical test of the assumption that a series of numbers, typically asset returns, represent a random sample from a normal distribution.

Sharpe Ratio. A signal-to-noise ratio based on classical standard deviation measurement of noise. It is calculated as the ratio of expected excess returns to the standard deviation.

Sortino Ratio. A signal-to-noise ratio based on downside deviation, a one-sided measurement of noise. It is the ratio of expected excess returns to the downside deviation. This is of interest when considering the distribution of asset returns, where downside scatter represents financial risk.

Downside deviation. A non-symmetric measure of variability in only values that lie below some measure of center or reference value.

Omega indicator. This indicator was proposed by Keating and Shadwick in 2002. It takes into account all the moments of the asset return distribution function in a very simple way.

Future Work

An initial overlay analysis was performed on datasets from May 2007 onward, using a variety of parameters and methods. No strong benefit was initially apparent and we remark on the danger of over fitting the signal with these alternatives. However, we have now entered into our signal and portfolio optimization phase of development. We feel there is much room for improvement over the current signal and trading scheme, and we expect higher returns for this portfolio system as a result.

Disclaimer

This summary report is not authorized for use as an offer of sale or a solicitation of an offer to purchase investments in any fund managed by Poulin Holdings or its affiliates. Offerings may be made only by means of a confidential private offering memorandum and are suitable only for eligible investors who are accredited investors and who are financially sophisticated.

Contact

Chris Poulin
P.O. Box 15664
Boston, MA 02215

Email:
chris@poulinhugin.com
Phone : +1 617 755 9049