Predictability of the daily high and low of the SP 500 index

Jones, Clive

Economic Data Resources LLC

1 March 2015

Predictability of the Daily High and Low of the S&P 500 Index

Clive Vaughan Jones

Economic Data Resources LLC

PO Box 1009

Boulder, Colorado

USA

Email: businessforecastblog@gmail.com

Tel. 303-440-7626

March 1 2015

Abstract

Ratios involving the current period opening price and the high or low price of the previous period are significant predictors of the current period high or low price for many stocks and stock indexes. This is illustrated with daily trading data from the S&P 500 index. Regressions specifying these “proximity variables” have higher explanatory and predictive power than benchmark autoregressive and “no change” models. This is shown with out-of-sample comparisons of MAPE, MSE, and the proportion of time models predict the correct direction or sign of change of daily high and low stock prices. In addition, predictive models incorporating these proximity variables show time varying effects over the study period, 2000 to February 2015. This time variation looks to be more than random and probably relates to investor risk preferences and changes in the general climate of investment risk.

1. Introduction

Ratios involving the current period opening price and the high or low price of previous periods are significant predictors of the current period high or low price for many stocks and stock indexes. The resulting models lead to out-of-sample predictions[1] which outperform autoregressive and "no change" models.

These points are illustrated with trading data from the S&P 500 index

Ratio or proximity variables are defined as differences between the opening price for the current period and the high or low of a preceding period, scaled by the high or low of this previous period. The target variable for predictive relationships in this study is growth in the daily high or low of the S&P 500 index.

Research of George and Hwang (2004) provide precedent for this type of model, when they suggest building portfolios with stocks whose prices are near their 52-week high as a way of generating superior returns. More recently, Li and Yu (2011) show that proximity to the 52-week high predicts market returns.

The focus here, however, is on quantitative models which predict at higher time frequencies – specifically, trading days.

Ordinary least squares (OLS) regressions specifying "proximity variables" achieve lower out-of-sample mean absolute percentage error (MAPE) and mean square error (MSE) than benchmark models over the study period, January 3, 2000 through February 25, 2015.

Similar performance is achieved by predictive models specifying proximity variables for daily lows of the S&P 500.

A study of regression coefficients also shows that coefficients of the ratio or proximity variables vary over time in a manner that may be associated with the investment climate and investor risk preferences.

The forecast performance of regressions with these proximity variables have higher explanatory and predictive power than is typical with

---

[1] When model parameters are estimated over one set of data, and another set of data is used to populate the values of explanatory variables to generate predictions, the procedure is called "out-of-sample" prediction. The term "pseudo out-of-sample" sometimes is encountered, but there seems to be no effective distinction, unless one wishes to employ "out-of-sample" as a strict synonym for real time forecasting.

forecasting relationships developed with stock prices or returns. These explanatory variables, furthermore, appear to be little discussed, if not novel, in the financial literature.

The organization of this paper is as follows. The following Section describes the data on the S&P 500 index. Section 3 presents specifics of OLS regressions with the trading-day-over-trading growth in daily high and low prices as independent variables, and ratio or proximity variables as explanatory variables. Section 4 discusses the performance of rolling regressions with these variables in out-of-sample prediction, comparing with "no change" and autoregressive model forecasts. MAPE and MSE values are provided, together with metrics of the performance of the proximity model (Pvar) regressions in predicting the direction of change of the daily high and low prices. The fifth Section reports findings relating to time variation of the coefficients of these Pvar regressions. Section 6 discusses significance of these findings, offering an interpretation in terms of heuristics for buying and selling stock, based on the spread between the opening price and the previous period high or low. Section 7 offers a summary and conclusion.

2. Data

Trading day data on the opening, high, and low prices for the S&P 500 index was obtained from Yahoo Finance (http://finance.yahoo.com/). The data span January 3, 1990 through February 25, 2015, a total of 6,337 trading days.

The Standard & Poor's 500 is a stock market index based on the market capitalizations of 500 large companies having common stock listed on the NYSE or NASDAQ. S&P 500 index components and their weightings are determined by S&P Dow Jones Indices, capturing approximately 80% coverage of available market capitalization.

3. Predictive Models for Daily High and Low Prices

Table 1 presents the results of linear regressions onto trading day growth in the high and low prices of the S&P 500 index estimated with daily trading data from 2000 to 2015.

Table 1
OLS Regressions on Growth of Daily High and Low Prices

	Daily High		Daily Low	
	Coefficient	StanError	Coefficient	StanError
OPH_t	0.882	*0.022*	0.256	*0.025*
OPL_t	0.195	*0.017*	0.881	*0.020*
OPH_{t-1}	-0.271	*0.017*	0.226	*0.020*
OPL_{t-1}	0.161	*0.014*	-0.210	*0.017*
Constant	0.002	*0.0002*	-0.002	*0.0002*
Observations	3810		3810	
R^2	0.487		0.504	
Durbin-Watson	2.100		2.080	

Coefficients, standard errors, and details of the ordinary least squares (OLS) regression for daily high prices are on the left of the table, while OLS results for daily low prices are shown on the right of Table 1.

The dependent or target variables are growth in the high price GH_t and growth in the low price GL_t of the S&P 500 over successive trading days. These variables are defined as,

$GH_t = (H_t - H_{t-1})/H_{t-1}$

$GL_t = (L_t - L_{t-1})/L_{t-1}$

where H_t and L_t refer to the daily high and low prices in trading day t. Here t ranges over calendar dates which can be mapped onto a series from 1 to 3810.

Four explanatory variables are specified in the linear regressions for the high and low prices. All of these independent variables are calculated with the current period opening price O_t and are defined as,

$OPH_t = (O_t – H_{t-1})/H_{t-1}$

$OPL_t = (O_t – L_{t-1})/L_{t-1}$

$OPH_{t-1} = (O_t – H_{t-2})/H_{t-2}$

$OPL_{t-1} = (O_t – L_{t-2})/L_{t-2}$

Significance of Results

For both the daily high and daily low regressions, standard errors (StanError) of coefficients and constant terms in Table 1 indicate statistical significance by conventional standards. Durbin-Watson statistics suggest acceptance of the null hypothesis of no serial correlation for both regressions, given 3810 degrees of freedom. The values of R^2 or the coefficient of determination are high for such financial returns or stock price regressions, comparing with R^2 of less than 0.02 for first order autoregressive relationships predicting the growth of the daily highs or lows of the S&P 500.

Dynamics

The coefficients in Table 1 suggest fairly complex dynamics. Thus, focusing on regression results for GH_t, if the current opening price O_t is greater than H_{t-1}, OPH_t and OPL_t both will be positive or greater than zero. Their contribution to the growth of the current period high price, therefore, will be positive. The coefficients of OPH_{t-1} and OPL_{t-1}, on the other hand, have opposite signs. It is then a matter of the specific numbers whether the net effect will be positive or negative.

For example, there is a case in which the current opening price is less than the high for two trading days previous, but above the low for that trading day. This would lead unambiguous growth in GH_t.

4. **Out-of-Sample Predictive Performance**

The real test of a stock market price prediction model is probably whether it performs out-of-sample (Welch and Goyal, 2008).

MAPE and MSE

What we can call "regressions with proximity variables" or Pvar achieve significantly lower MAPE's and MSE's than benchmark models, such as the no-change forecast and forecasts based on first order autocorrelation regressions.

These comparisons are shown in Table 2.

MAPE's and MSE estimates are based on predictions from rolling regressions developed with five years of history.

Out-of-sample predictions from the Pvar regressions and autoregressions start with January 3, 2000.

The details are as follows. Four OLS regressions are developed for each trading day from 1:3:2000 through February 25, 2015. Two regressions specify OPH_t, OPL_t, OPH_{t-1}, OPL_{t-1} as the explanatory variables, and GH_t

and GL_t as the dependent or target variables. Two others autoregressions map G_{t-1} onto G_t and L_{t-1} onto L_t.

The "no change" forecast, of course, involves identifying the high or low for the previous trading day, and using them as predictions for the current trading day.

Results from rolling regressions specifying the explanatory variables OPH_t, OPL_t, OPH_{t-1}, OPL_{t-1} are listed under "Pvar regression."

The label "AR estimate" in Table 2 refers to autoregressions of OPH_{t-1} onto OPH_t or OPL_{t-1} onto OPL_t.

Instances when forecasts from these rolling regressions are not strictly out-of-sample are removed from the data used in calculating these error metrics. Thus, the opening price equals the daily high or low 1186 times in this more than fourteen year period. Absolute percent and squared errors for these trading days are removed from calculations contributing to Table 2 without changing, it can be noted, the ranking of the error metrics.

Table 2

MAPE and MSE of OLS Regressions with Proximity Variables and Benchmark Forecasts

Forecasts for Daily Highs						
MAPE				MSE		
Pvar regression	AR estimate	No Change		Pvar regression	AR estimate	No Change
0.37%	0.58%	0.57%		39.4	103.8	99.1
Forecasts for Daily Lows						
MAPE				MSE		
Pvar regression	AR estimate	No Change		Pvar regression	AR estimate	No Change
0.43%	0.70%	0.68%	0.57%	54.7	144.8	139.1

The Pvar regressions specifying the proximity variables achieve dominance in predictive performance.

For example, the MAPE's for Pvar forecasts for the daily high are 65 percent lower than MAPE's of the benchmark forecasts. Similarly, the mean square errors (MSE's) of the no change and AR forecasts of the daily high are more than 2.5 times higher than the MSE of the OLS regression with proximity variables. It is interesting that the no change forecasts generally produce lower errors by both metrics than the AR estimates.

Predictions of the Direction of Change of Daily Highs

The regressions with proximity variables also achieve large gains in forecast accuracy over benchmark models in predicting the direction of change of daily highs and lows.

This involves counting instances in which the signs of the predicted growth rates for the daily highs and lows equal the signs of the actual growth rates for these variables in a trading day.

The rolling regressions for predicting daily highs predict the correct sign of the trading-day-over-trading day growth in the high of S&P 500 index 78.2 percent of the time, over the period 1:3:2000 to 2:25:2015. The first order autoregressive model for the daily highs, by comparison, correctly predicts the direction of change of the daily high only 51.7 percent of the time in this period. The regression forecast models for the daily lows predict the direction of change of the daily low 77.6 percent of the time. These proportions are based on the more than 2600 predictions which can be said to be truly out-of-sample.

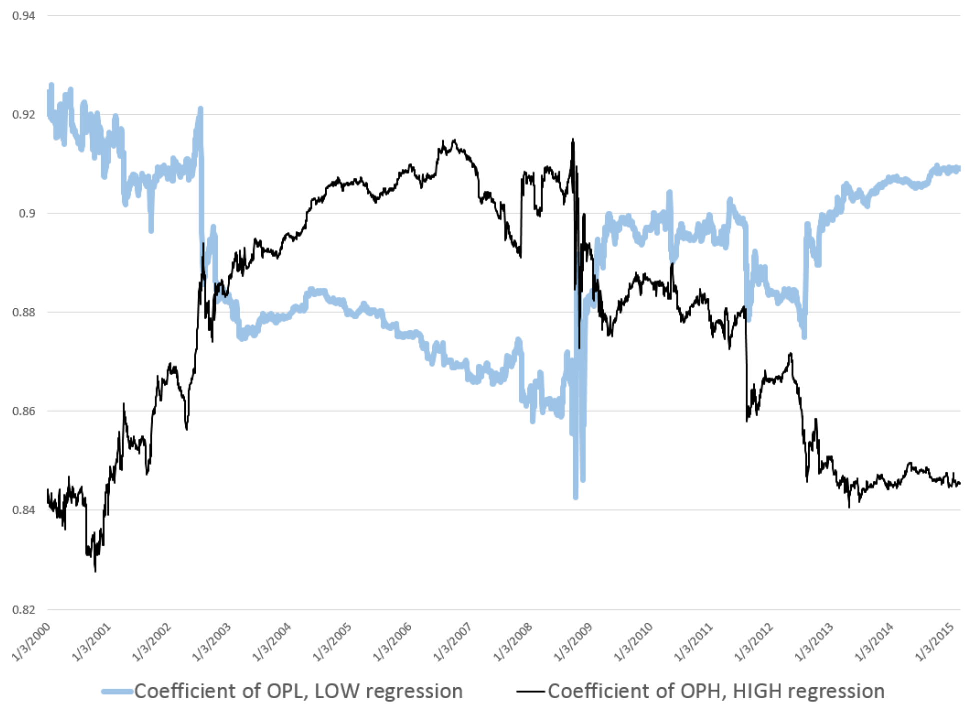
5. Time Varying Coefficients

One issue highlighted in the literature on the predictability of stock prices is whether coefficients of predictive relationships vary over time (Lo, 2005).

In this regard, there is evidence for time varying coefficients for the "proximity variables," perhaps signifying investor responses to time-varying risk.

Thus, Figure 1 charts the coefficients associated with OPH_t and OPL_t over the rolling regressions described in the preceding section, each estimated with five years of history on trading day data.

Figure 1 Evidence for Time Varying Coefficients - Estimated Coefficients of OPH_t and OPL_t Over Study Sample



There are abrupt changes in the values of the coefficients of OPH_t and OPL_t in 2008. These plausibly reflect stock market volatility in the Great Recession. After 2010 the value of both coefficients tends to move back to levels seen at the beginning of the study period.

This suggests trajectories influenced by the general climate of risk for investors and their risk preferences – an interpretation expanded in the next following section.

Consideration of the time variation of these coefficients also has implications for out-of-sample forecast errors. Thus, late 2008, when values of the coefficients of both OPH and OPL make almost vertical movements in opposite directions, is the period of maximum out-of-sample forecast errors. Forecast errors for daily highs, for example, reach a maximum of 8 percent in October 2008. This can be compared with typical errors of less than 0.4 percent for out-of-sample forecasts of daily highs with the proximity variable regressions.

6. Interpretation and Significance of Findings

The Table 1 regressions and the forecast performance of updated regressions with these specifications seem to have higher explanatory and predictive power than is typical with forecasting relationships developed with stock prices or returns. These explanatory variables, furthermore, appear to be little discussed, if not novel, in the financial literature.

There is increasing evidence for predictability of stock market prices and returns (Lim and Brooks, 2011), with predictability being more salient during some periods than others.

Rapach and Zhou (2013) demonstrate out-of-sample predictability for stock returns with ensemble and factor methods. They comment that predictability in this context is consistent with efficient markets, if predictive models capture exposure to time-varying aggregate risk.

At the same time, investors may follow heuristics such as "buy when the opening price is greater than the previous period high" or "sell, if the opening price is lower than the previous period low." This is suggested by the time trajectory of estimated regression coefficients shown in Figure 1.

Recall, for example, that the coefficient of OPH_t measures the influence of the spread between the opening price and the previous period high on the growth in the daily high price. The trajectory, shown in the narrow, black line, trends up in the approach to 2007. This may reflect investors' greater inclination to buy the underlying stocks, when the opening price is above the previous period high. But then the market experiences the crisis of 2008, and investors abruptly back off from their eagerness to respond to this "buy" signal. With onset of the Great Recession, investors become increasingly risk adverse to such "buy" signals, only starting to recover their nerve after 2013.

A parallel interpretation of the trajectory of the coefficient of OPL_t can be developed based on developments 2008-2009.

If this interpretation is correct, time variation in these coefficients is more than random and bears relationship to investor risk preferences and the general climate of investment risk.

Note that confidence intervals for forecasts from these regressions involve advanced computation, since the residuals are *non-Gaussian, sharp peaked and heavy tailed* – which, as Cont (2001) notes, often is true for stock and financial returns.

Major findings of this research, however, do not depend on confidence intervals, the focus being on what one might call out-of-sample performance metrics.

Conclusion

Ratios involving the current period opening price and the high or low price of the previous period are significant predictors of the current period high or low price for many stocks and stock indexes. This is illustrated here with daily trading data from the S&P 500 index. Regressions specifying these variables have higher explanatory and predictive power than benchmark autoregressive and "no change" models. These "proximity variables," furthermore, appear to be little discussed, if not novel, in the financial and econometric literature.

The proximity variable or Pvar regressions show lower MAPE and MSE for the study period dating from the first trading day in 2000, and, additionally, correctly predict the direction of change of daily high and low prices roughly 80 percent of the time.

An in-depth look at the coefficients for the more than 3,800 rolling Pvar regressions for daily highs in the study period provide evidence for time-variation. The coefficients of OPH_t and OPL_t show sharp movements in 2008, plausibly related to financial dislocation associated with the Great Recession. October 2008 appears to be especially significant, since the closing prices of S&P 500 and coefficients of OPH_t and OPL_t make almost vertical movements in opposite directions that month.

Thus, time variation in the coefficients of these proximity variables looks to be more than random and probably relates to investor risk preferences and changes in the general climate of investment risk.

Avenues for further research include extension of these relationships across groupings of trading days and exploring similar relationships with other stocks and stock indexes in the United States and internationally.

References

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues, Quantitative Finance, Volume 1, Issue 2, 223-236.

George, T. J., Hwang, C. Y. (2004). The 52-Week high and momentum investing. Journal of Finance 59, 2145-2176.

Li, J., Yu, J. (2012). Investor attention, psychological anchors, and stock return predictability, Journal of Financial Economics, 104, Issue 2, 401–419.

Lim, K., Brooks, R. (2011). The Evolution of Stock Market Efficiency Over Time: A Survey of the Empirical Literature, Journal of Economic Surveys, 25, No. 1, 69–108.

Lo, A., (2005). Reconciling Efficient Markets with Behavioral Finance: The Adaptive Markets Hypothesis, Journal of Investment Consulting, Vol. 7, No. 2, 21-44.

Rapach, D., Zhou, G. (2013). Chapter 6 - Forecasting Stock Returns, Handbook of Economic Forecasting, Volume 2, Part A, 2013, 328-383.

Welch, I., Goyal, A. (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction, Review of Financial Studies, 21, Number 4, 1455-1508.