A Nonlinear Approach for Predicting Stock Returns and Volatility with the Use of Investor Sentiment Indices

Stelios Bekiros
European University Institute

Rangan Gupta
University of Pretoria

Clement Kyei
University of Pretoria

June 2015
A NONLINEAR APPROACH FOR PREDICTING STOCK RETURNS AND VOLATILITY WITH THE USE OF INVESTOR SENTIMENT INDICES

STELIOS BEKIROS a,b *, RANGAN GUPTA a,c ** and CLEMENT KYEI c ***

a IPAG Business School, b European University Institute (EUI), c University of Pretoria

ABSTRACT

The sentiment-based investor indices $S^{BW}$ and $S^{PLS}$ introduced by Baker and Wurgler (2006, 2007) and Huang et al. (2015) respectively are commonly reported to predict monthly stock returns, yet based on a linear causality framework. However, the latter may lead to misspecification and lack of robustness. We provide statistical evidence that the relationship between stock returns, $S^{BW}$ and $S^{PLS}$ is characterized by structural instability and inherent nonlinearity. Moreover, using a nonparametric causality approach, we show that $S^{BW}$ and not $S^{PLS}$ is in fact a stronger predictor of market returns and volatility, as opposed to previous empirical evidence.

JEL Codes: C22, C32, C53, G10, G11

Keywords: Investor sentiment; stock markets; nonlinear dependence

1. INTRODUCTION

As pointed out by Huang et al. (2015), investor sentiment can affect asset prices as market agents tend to make overly optimistic or pessimistic judgments and choices. However, trader behaviour and investor sentiment is not directly measurable or observable. Based on that fact, Baker and Wurgler (2006, 2007; hereafter BW) constructed a novel sentiment index, which aggregates information from six proxies namely close-end fund discount rates, share turnover, number of initial public offerings (IPOs), first-day returns of IPOs, dividend premium and equity share in new issues. The Baker and Wurgler (2006, 2007) index is created via a principal components methodology. Recently, Huang et al. (2015) using the same proxies of BW developed a new “aligned investor sentiment index”, which separates out the information

* Corresponding author: a IPAG Business School, 184 Boulevard Saint-Germain, 75006 Paris, France.; Tel.: +33 01 53 63 36 00 ; Fax: +33 01 45 44 40 46 ; b European University Institute (EUI), Via della Piazzuola; 43, I-50133, Florence, Italy; Tel.: +39 055 4685 916; Fax: +39 055 4685 902; E-mail address: stelios.bekiros@eui.eu.

**, Department of Economics, University of Pretoria, Pretoria, 0002, South Africa. E-mail address: rangan.gupta@up.ac.za.

*** Department of Economics, University of Pretoria, Pretoria, 0002, South Africa. E-mail address: kweku.shaker@gmail.com.
in the proxies that is relevant to the expected stock returns from the error or noise, utilizing a partial least squares (PLS) method. Importantly, Huang et al. (2015) showed that their modified index can predict aggregate stock market returns at a monthly frequency, as opposed to the results by Baker and Wurgler (2007) and Baker et al. (2012).

The objective of our work is to compare the predictive ability of the Baker and Wurgler (2006, 2007) sentiment index ($S^{BW}$) against that of Huang et al. (2015) (denoted as $S^{PLS}$) not only for the aggregate stock market returns, but also for their volatility, using the nonparametric causality test of Nishiyama et al. (2011). This test is developed to incorporate higher-order interrelationships inherently based on the nonlinear dependence structure between the investigated variables in question. Our assessment to use a nonparametric approach also stems from the fact that the behaviour of stock returns and volatility vis-à-vis $S^{BW}$ and $S^{PLS}$ might be highly nonlinear.

It is worth noting that Huang et al. (2015) obtained their results on the predictability of stock returns via the $S^{PLS}$ index, based on a linear regression framework. Whilst this is undoubtedly most widely used in the literature (e.g., Rapach and Zhou, 2013), yet if statistical testing reveals nonlinearity - the existence of which we demonstrate thereafter – the results are unreliable. This is the first paper, to the best of our knowledge, which utilizes a nonlinear nonparametric framework to compare the forecastability of $S^{BW}$ and $S^{PLS}$ vis-à-vis market returns and their volatility, the latter of which has never been examined whatsoever. The rest of the paper is organized as follows: Section 2 presents the empirical methodology, while Section 3 discusses the data and presents the results. Finally, Section 4 concludes.

2. METHODOLOGY

The methodology by Nishiyama et al. (2011) explores high-order causalities assuming the following nonlinear dependence between the investigated series:

$$x_t = g(x_{t-1}) + \sigma(y_{t-1})\epsilon_t$$ (1)
where \( \{x_t\} \) and \( \{y_t\} \) are stationary processes and \( g(\cdot) \) and \( \sigma(\cdot) \) are unknown functions which satisfy certain conditions for stationary. In general, \( y_{t-1} \) carries information in predicting \( x^K_t \) for a given \( K \), hence the null hypothesis of non-causality in the \( K \)th moment is given by:

\[
H_0: \mathbb{E}(x^K_t|x_{t-1}, \ldots, x_1, y_{t-1}, \ldots, y_1) = \mathbb{E}(x^K_t|x_{t-1}, \ldots, x_1) \text{ w.p. } 1. \tag{2}
\]

where \( \text{w.p. } 1 \) abbreviates to "with probability one". Formally, \( y_t \) does not cause \( x_t \) up to the \( K \)th moment if:

\[
H_0: \mathbb{E}(x^K_t|x_{t-1}, \ldots, x_1, y_{t-1}, \ldots, y_1) = \mathbb{E}(x^K_t|x_{t-1}, \ldots, x_1) \text{ w.p. } 1. \text{ for all } k = 1, \ldots, K \tag{3}
\]

Nishiyama et al. (2011) in their work provide a detailed description of how to construct the test statistic \( S_t^{(k)} \) for any \( k = 1, \ldots, K \). We implement the test for \( k = 1 \) to test for non-causality in the 1st moment (mean) and for \( k = 2 \) in the 2nd moment, i.e., non-causality in variance.

### 3. Empirical Analysis

The aggregate stock market returns are estimated as the excess returns of a market index, which is common in the relevant literature. Specifically we calculate the continuously compounded log-returns of the S&P 500 index (including dividends) minus the risk-free rate. The return \( (spr) \) of the S&P 500 and its volatility \( (spv) \) measured as the squared values of the returns, are derived from the Center for Research in Security Prices (CRSP). The data on both the value-adjusted CSRP for the S&P500 index and the risk free rate can be found at http://www.hec.unil.ch/agoyal/, while the sentiment indexes \( S^{BW} \) and \( S^{PLS} \) can be downloaded from http://apps.olin.wustl.edu/faculty/zhou/Sentiment_BW_PLS.txt, already standardized. Our monthly sample covers the period 1965:7 - 2010:12.

Initially we employ the standard linear Granger causality test for complementarity reasons. To ensure that our results will be comparable vs. the nonparametric test, we use a lag-length of one in the vector autoregressive (VAR) modelling, as selected by the SIC criterion. As it is demonstrated from
Table 1, the null hypothesis that $S^{BW}$ and $S^{PLS}$ does not Granger cause $spr$ can be strongly rejected only for the latter case, thus confirming the results of Huang et al. (2015).

[Please insert Table 1]

Next, we conduct the Bai and Perron (2003) test for detecting multiple structural breaks in case of an autoregressive AR(1) model for $spr$, as well as for a bivariate VAR(1) including $S^{BW}$ or $S^{PLS}$ respectively. We were not able to detect any structural breaks for the AR(1) model, yet interestingly five breaks were obtained by the VAR model between $spr$ and $S^{BW}$ (i.e., 1970:7, 1974:10, 1977:1, 2000:9 and 2008:9), and between $spr$ and $S^{PLS}$ (i.e., 1972:6, 1974:10, 1987:10, 1991:3 and 2008:10). In the presence of these breaks, the assumption of parameter constancy over the entire sample as inferred by the linear Granger causality test is strongly violated, and consequently cannot be deemed conclusive.

[Please insert Table 2]

Furthermore, we use the Brock et al. (1996) test for the non-iid null hypothesis on the residuals of the $spr$ AR(1) model and of the VAR comprising $spr$ and $S^{BW}$ or $S^{PLS}$ respectively. It is illustrated from Table 2, that the test overwhelmingly rejects the null of iid structure for many embedding dimensions, thus implying an omitted nonlinear structure. In order to substantiate the evidence of structural breakpoints, nonstationarity and possible nonlinear interdependencies, and eventually deal with the misspecification of our linear modelling, we utilize the nonparametric causality test proposed by Nishiyama et al. (2011). The results are displayed in Table 3. We observe that in accordance with the Granger causality outcome, (i) the $S^{PLS}$ operates as a predictor for $spr$ and (ii) $S^{BW}$ causes $spr$. Those results also carry over to the volatility series ($spv$) as well. Nevertheless, it is impressive that the nonparametric testing tends to suggest that $S^{BW}$ is a stronger predictor compared to $S^{PLS}$ both for returns and volatility. Apparently, based on the findings of nonlinear structural relationships between stock returns and the two investor sentiment indices, we deem the detected causality results for $S^{BW}$ and $S^{PLS}$ obtained by the Nishiyama test, far more robust and reliable.

[Please insert Table 3]
4. Conclusions

As market sentiment is not directly observable, Baker and Wurgler (2006, 2007) created a novel investor index ($S^{BW}$) that aggregates behavioural information from six financial proxy indicators using principal components analysis. More recently, Huang et al. (2015) using the same proxies developed a new index ($S^{PLS}$) which distinguishes information from the observed expected stock returns vis-à-vis error or noise signals in the market, via a partial least squares approach. Based on a linear predictive regression model, they showed that their modified index could predict aggregate stock market returns at monthly frequencies.

In the present work, we showed that due to inherent nonlinearities in the interrelationship between returns and the two sentiment indexes, the linear Granger causality framework upon which Huang et al. (2015) relied, might lead to misspecification. In this regard, we employed the nonparametric causality test of Nishiyama et al. (2011) to demonstrate that a new directionality from $S^{BW}$ to stock returns and volatility emerged as opposed to previous empirical evidence. More importantly, we managed to substantiate that $S^{BW}$ is a stronger predictor of market returns and volatility compared to $S^{PLS}$. In future research, it would be interesting to verify whether our results are robust to out-of-sample forecasting or not, using various nonlinear models.
REFERENCES


**TABLE 1: LINEAR GRANGER-CAUSALITY TEST**

Dependent Variable: \( spr \) (1965:7-2010:12)

| \( S^{BW} \) | 1.4043 |
| \( S^{PLS} \) | 8.6395*** |

Note: *** indicates rejection of the null hypothesis of absence of Granger causality at 1% level.

**TABLE 2: BDS TEST**

<table>
<thead>
<tr>
<th>Dimension</th>
<th>AR(1)</th>
<th>( S^{BW} )- based VAR(1)</th>
<th>( S^{PLS} )- based VAR(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.0049</td>
<td>0.0055</td>
<td>0.0095</td>
</tr>
<tr>
<td>3</td>
<td>0.0007</td>
<td>0.0006</td>
<td>0.0009</td>
</tr>
<tr>
<td>4</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>5</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>6</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Note: Entries are \( p \)-values for the null of serial independence in the error structure of \( spr \) after using an AR(1) filter or two VAR(1) model specifications i.e., \([ spr, S^{BW} ]\) and \([ spr, S^{PLS} ]\) respectively.

**TABLE 3: NONPARAMETRIC CAUSALITY TEST**

Dependent Variable: \( spr \) and \( spv \) (1965:7-2010:12)

<table>
<thead>
<tr>
<th>Test statistics</th>
<th>( S_T^{(1)} )</th>
<th>( S_T^{(2)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S^{BW} )</td>
<td>153.58**</td>
<td>149.40**</td>
</tr>
<tr>
<td>( S^{PLS} )</td>
<td>143.16**</td>
<td>90.78 **</td>
</tr>
</tbody>
</table>

Note: ** indicates rejection of null hypothesis of non-causality at a 5% level (14.38); \( S_T^{(1)} \): Test statistic for causality in-mean; \( S_T^{(2)} \): Test statistic for causality in-variance.