Investor Psychological Bias and Speculation: Asymmetric Impacts of Big Data on Commodity Price

Research-in-Progress

Xin Li
School of Management, University of Chinese Academy of Sciences
No. 55, Zhongguancun East Road, Haidian District, Beijing, China
leexin111@163.com

Abstract

This research attempts to propose a new measure of investor psychological bias with big data crawled from the web. The author constructs the investor bias measure with web search data and investigates the influences of this new measure on crude oil futures prices. Using 225,250 data points from Google, this paper computes the investor bias, and evaluates it with trading volumes. The author establishes a Markov switching model and reveals that the influences of investor bias on crude oil prices are asymmetric in two regimes (rising and decreasing) of crude oil prices. The influence of the new measure on oil prices is negative (positive) in the phases of prices decreasing (rising). This study contributes the new measurement of investor psychological bias leveraging big data crawled from the web to the research community. Big data may become an important source to better understand investors’ trading psychology, and support their decision making.

Keywords: Psychological bias, speculation, commodity price, asymmetry, big data analytics

Introduction

In financial market, it has been demonstrated that investor sentiment has played an important role in affecting commodity price (Antweiler and Frank 2004, Oh and Sheng 2011). Some previous studies suggest that investor sentiment during recession phase tend be more intense than the sentiment during the period of prosperity (Tetlock 2007, Garcia 2013). This asymmetric attention on market should possibly result in different changes of commodity price during good and bad times. Evidence from behavioral finance presents that prices may deviate from fundamental values since individuals behave irrationally. Investor psychological bias such as overconfidence and limited attention, as a kind of investor sentiment, reflects that investors may be irrational. For example, individual investors tend to exhibit loss-averse behavior due to limited attention. Overconfident traders may contribute to speculation in financial market (DeBondt and Thaler 1985). Empirical evidences in financial security show that bias from individual investors, institutional investors, and even analysts will affect prices (Daniel et al. 2002). Financial commodity prices may be affected by investor bias. Speculative trading represents investors’ psychological sentiment like positive, negative or neutral bias, which may result in fluctuations of commodity prices. Therefore, it is useful to investigate the relationship between investor psychological bias and commodity prices so as to explicitly depict speculators’ influences and instruct the decision making of investors in financial commodity markets.

However, due to the lack of rapidly generated data sets describing the investor psychological behavior, it is hence difficult to timely and accurately measure investor attention or sentiment. Previous studies focused on news, unusual trading volume, and extreme returns to measure investor attention (Barber and Odean 2008). However, all these proxies just reflect investors’ psychological bias indirectly. For example,
if some stocks appear in news report, we can judge that investors may pay attention to those stocks and buy them. Compared to these existing proxies, measures constructed from the web are possibly a better choice. Internet access can increase trading volumes, so Internet provides vast data sources to construct new measures for investor psychological bias (see Choi et al. 2010). Accordingly, I establish new direct proxies using the extracted user generated content (UGC) from Internet to represent investors' behavior before investors really start to trade in markets. UGC can reflect trading motivation of investors, especially speculators since investors tend to use different kinds of UGC sources to show their attention. Then I use new measures to investigate the contribution of investor bias to commodity price. Furthermore, I attempt to interpret how web search data shapes the relationship between investor psychological bias and commodity price.

The motivation of this work is to give a clear cognition about investor psychological bias and commodity price with the adoption of big data. I collected web search data series and constructed investor psychological bias accordingly. And more importantly, to identify the influences of investor bias on crude oil prices can better support the decision making of market participants. This work selects crude oil market as the research domain since crude oil has a significant role in global financial market and has drawn great attention from investors including institutional and individual investors. It is meaningful to explore the impacts of psychological bias, since complicated factors including supply and demand, stock prices, and market expectations affect crude oil prices. Other financial commodities like gold could also be investigated by focusing on investor psychological trading behavior. The major focus of this work is to construct a new measure of investor psychological bias using big data from the web and investigate whether it affects crude oil prices in an asymmetric way.

In this paper, I have conducted the following experiments and analysis in this work. First, I construct a new measure for investor psychological bias from web search data. Second, I establish a Markov switching model and estimate the influences of this bias and other independent variables like oil inventory and S&P 500 Index. Third, I discuss how the new measure captures investors’ attention. Following the empirical studies, I have proved that the measure from web search has an obvious and asymmetric influence on crude oil prices in phases of oil rising and decreasing.

The rest of the paper is organized as follows. Section II provides a background of behavior economics, existing measures, and crude oil prices. Section III describes the data and methodology. Section IV presents empirical analysis. Section V discusses implications and contribution. Section VI concludes.

**Literature Review**

Relevant literature is reviewed in this section to present the existing work. First, some studies about investor psychological bias are shown to introduce some basic concepts of behavior economics. Second, existing measures of investor bias are reviewed. Third, literature on crude oil prices fluctuation is comprehensively surveyed.

**Investor Psychological Bias**

To present how speculators make decisions about trading positions or futures contracts, I have reviewed previous classical literature discussing investor behavior bias in order to interpret speculation from the most fundamental economic theory of investor psychology. Some economists such as Adam Smith, Irving Fisher and John Maynard Keynes argued that imperfect rationality affects investment decisions and market outcomes. Individuals make decisions by maximizing a utility function which is assumed to be time-consistent and depend on framing and reference points (Baucells and Heukamp 2006, Baucells, Weber and Welfens 2011).

In the existing literature, DellaVigna (2007) discussed time preferences, risk preferences, and social preferences in economic fields. They have documented how consumers deviate from the standard model in the choices of credit card, housing prices, insurance contracts, and loans where the agents show nonstandard preferences and beliefs in the markets, and provided evidences to understand markets and institutions so as to explain individual behavior. DeLong (1990) analyzed the factors of arbitrage, and considered that mispricing was stochastic, persistent, and correlated with noise traders. If arbitrageurs are risk averse and have a limited investment horizon, the noise traders affect the equilibrium price in
spite of arbitrage. If noise traders are bullish about market situations, they will bid the bubble price higher. But the arbitrageurs do not know whether the mispricing will get even worse in the next period, they cannot short the shares aggressively enough if they are given the short time horizons. They also showed that the noise traders who would like to take more risks outperformed the rational traders under some situations. Other detailed survey could be seen in (Lucey and Dowling 2005, Kliger and Kudryavtsev 2008, Tourani-Rad and Kirkby 2005, Glaser et al. 2007, Shapira and Venezia 2001).

Over-confidence, limited attention and familiarity bias could cause the formation of nonstandard behavior and influence the price (Daniel, Hirshleifer and Teoh 2002). When investors chose stocks or futures investment, they tend to prefer the ones that they are familiar with. The behavior is defined as familiarity bias in Fox and Tversky (1995). Existing evidences indicate imperfect rationality affects trading, expectations and prices in capital markets. Psychological bias is relevant with heuristic simplification and emotional judgment, and explain salience and availability effects. Due to limited attention and mental processing power, individual investors show loss-averse behavior. Individuals are concerned about gains and losses as measured relative to an arbitrary reference point, and they are more prone to realizing gains than losses. Specifically, stocks investors chose to sell subsequently outperform the ones that investors retain. The individual investor behavior that Odean observes (Barber and Odean 2008): keep your winners and sell your losers, which could be considered as an indication of individual biases. In contrast with individual investors, institutional investors like mutual funds tend to buy high momentum stocks and sell low momentum stocks. Investors tend to use past performance to instruct their future investing decisions, which is especially true of mutual funds and stock purchase decisions. It has been proved that investor biases affect security prices substantially. Some indications show that there is much misallocation of resources in the economy as result of mispricing, which should be noted that imperfect rationality could cause damage needed regulation. Coval and Shumway (2005) provided strong evidences for behavior biases among investors and investigated the effect of these biases have on prices. These investors appear highly loss-average and have important short-term consequences for afternoon prices, as losing traders actively buy contracts at higher prices and sell contracts at lower prices than those prevailed previously. Their tests offer special advantages inevaluating behavior biases and any consequences they might have for price. Hirshleifer (2001) survey the investor psychology and asset pricing by documenting judgement and decision biases, risk, mispricing effect, and asset pricing theories based on investor psychology. Hirshleifer, Subrahmanyam and Titman (2006) proposed a model that irrational investors’ trade based on considerations that are not inherently related to fundamentals. Even stocks prices follow a random walk, irrational investors could influence cash flows in some situations. The new model include four important elements to illustrate the behavior of irrational investors: feedback from stock prices to future cash flows, irrational investors, time factor, and information about future cash flows. Hirshleifer and Teoh (2003) examined the consequences of limited attention for disclosure, financial reporting policy and market trading. They demonstrated that owing to limited attention, choices could affect investor perceptions and market price.

The existing researches indicate that investors can be irrational, over-confident, and have limited attention, all of which could possibly result in the extreme changes of price. Therefore, it is necessary to examine whether the speculators’ trading activities cause the surge and collapse of crude oil price, which is an important domain but there are not enough studies to explicitly highlight it.

Existing Measures of Investor Psychological Bias

I reviewed a few widely used existing measures of investor psychological bias in previous financial studies. The first is trading volume. Investors trading is relevant with their different expectation on markets. Hoffmann et al. (2010) also showed that overconfident traders tend to trade more aggressively since they believed they had more enrich information to support their decision. The second proxy is annual turnover, which reflects that investors are confident at their stocks (Barber and Odean 2001). Third, average return of securities sold can reflect overconfidence (Odean 1999). They also proved that stocks or commodities that had special performance were likely to chosen by investors. So absolute abnormal return ration is used as a measure of familiarity bias. Fourth, analyst coverage is another proxy for familiarity bias. Since more analyst covers a stock, more likely that it will grab attention of investors. More detailed information about these investor psychological bias can also be seen in Chen et al. (2007), Acker and Duck (2008), Graham et al. (2009), Grinblatt and Keloharju (2009), and Hoffmann et al. (2010).
This literature confirms that existing investor psychological measures are obtained from trading activities or securities performance. However, few proxies constructed using web-based data is used. With this, I next reviewed factors influencing crude oil prices.

**Crude Oil Prices Fluctuations**

Crude oil prices are deeply affected by supply and demand. The fluctuations of crude oil prices are related to fundamental value of crude oil in itself, the behaviors and psychological factors of heterogeneous agents in the market, and competition between agents (speculators, consumers and producers of crude oil) who have different personal utilities, beliefs and expectations (Hamilton JD (2009a). In financial commodity market, investors are identified as individual and institutional ones according to their different trading motivation. Institutional investors like mutual funds use futures market to hedge. Individual traders such as retailers, however, tend to speculate in futures market (Da et al. 2010). Some scholars recognized some individuals' extreme trading activities as a kind of bias that influences crude oil prices. For example, Woodard and Zhou (2009), Cifarelli and Paladino (2010), and Kilian and Murphy (2013) identified the influences of individual’s speculative trading on crude oil prices. Their models documented that the influences of speculation on crude oil prices were symmetric. However, another scholars including Einloth (2009) and Irwin et al. (2009) still held that supply and demand factors are the dominant ones affecting crude oil prices. Irrational trading behavior like speculation did not contribute to the volatility of crude oil prices.

Overall, these previous studies found mixed evidences on the role of speculative trading in oil prices by indirectly measuring this speculative trading activity. But these studies did not put forward the behavioral logics when they decide to change their positions or contracts in order to speculate. Therefore, it is of great importance to combine theories of behavioral economics with traditional empirical analysis based on econometric modelling.

**Data and Methodology**

**Data**

To depict investor psychological bias and its influence on crude oil futures price, I obtained data from four sources. First, the data set for investor bias construction was crawled from Google. Search data from Google represents millions of internet users using Google search engine to show their attention on financial market. One advantage of this web search data is that it is a kind of aggregated data source, which is the mutual reflection of different traders in the U.S. financial market.

Second, apart from aggregated web search data, I obtained 7,177,555 disaggregated data points (from June 2006 to February 2014) reflecting traders’ trading volumes from U.S. Commodity Futures Trading Commission in many commodity futures market. The data is separated into the following categories: producer/merchant/processor/user, swap dealers, managed money, and other reportable traders. I extracted trading positions of crude oil in New York Mercantile Exchange (NYMEX), filtered out other commodities, and finally obtained 74,210 points representing investors’ trading volume. I extracted these data to discuss the relationship among the new bias proxy and different trading volumes. Detailed data description is not listed in Table 1 because of a lack of space. Evaluation will be discussed in the empirical study.

Third, crude oil futures prices trading at NYMEX and oil stocks data are obtained from US Energy Information Administration.

Fourth, to obtain another investor psychological bias, I crawled internet forum data from the web and compare it with that from web search. The empirical work using these kind of data is in progress now, so the current empirical study mainly discusses the influence of bias from web search on crude oil futures prices.

**Methodology**

To construct investor psychological bias using web search data, I selected search terms that investors may use. These terms should cover aspects from oil supply, demand, production, and market news etc., all of
which can influence crude oil prices. I obtained 600 search terms about crude oil prices that are used on
the Internet use with the help of Google Correlate (Mohanna et al. 2011). These terms include ‘crude oil
and ‘oil nymex’ etc. After manually eliminating some irrelevant terms, 425 terms are finally obtained.
Using these terms, I extracted 225250 data points ranging from January 2004 to February 2014.

After extracting web search data series, I computed their Pearson correlation coefficients. The correlation
results show that many data series are strongly correlated with each other. Some data series with weak
correlation are eliminated. In addition, to solve their multicollinearity, I adopted principle component
analysis to obtain their common factors. The theoretical basis that web search can construct investor bias
proxy lies in that search is an active behavior to show investors’ attention, and may instruct their trading
activities. More search possibly indicates overconfidence and investors’ familiarity to some degree.

Accordingly, investor psychological bias is constructed using extracted web search data series:

\[
agg_{t}^{\text{agg}} = \alpha_{t}^{\text{sto}} + \alpha_{t}^{\text{pro}} + \alpha_{t}^{\text{futures}} + \alpha_{t}^{\text{oil}} + \alpha_{t}^{\text{nymex}} - \text{agg}_{t-4}^{\text{average}}
\]

Where \( t \) indicates weeks in crude oil prices, \( \text{bias} \) represent the new constructed metric, \( \text{sto} \),
\( \text{pro} \), \( \text{futures} \), \( \text{oil} \), \( \text{nymex} \) mean five normalized web search data series, which are ‘oil stocks’, ‘oil
production’, ‘crude oil price’, and ‘oil nymex’, respectively, and \( \text{agg}_{t} \), \( \text{agg}_{t-4}^{\text{average}} \) are aggregated search data and average search
values in past four weeks. All of these data series have highest correlation with real crude oil prices. Using this formula, I can capture the abnormal changes of aggregated web search data series. To eliminate the influences of dimension, I adopted normalization process using Z-score, which means the mean value is 0 and standard derivation is 1. Summary statistics of these data series are given at Table 1.

<table>
<thead>
<tr>
<th>Table 1. Summary Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td>Crude oil futures prices</td>
</tr>
<tr>
<td>(dollars per barrel)</td>
</tr>
<tr>
<td>Investor bias from web</td>
</tr>
<tr>
<td>search (normalized)</td>
</tr>
<tr>
<td>Oil inventories (normalized</td>
</tr>
<tr>
<td>barrel)</td>
</tr>
<tr>
<td>S&amp;P 500 Index (normalized)</td>
</tr>
</tbody>
</table>

**Empirical Analysis**

**Econometric Modelling**

The weekly West Texas Intermediate (WTI) futures prices start from January of 2004 through February
of 2014 at weekly frequency. I adopt a Markov switching regression to depict the patterns of crude oil
futures prices. Markov switching model has some unique advantages. First, it can specify the two regimes
(oil price rises and decreases) in oil futures price. Second, we can observe the different coefficients of
independent variables by estimating this model, and thus the potential asymmetry can be presented.
Third, the smoothed probability of this model will show the accuracy of this empirical test.

Considering a two-regime Markov switching process, where \( r_{t} = 0 \) means the rising phase, \( r_{t} = 0 \) means
the decreasing phase, and \( t \) means time. Accordingly, the transition probabilities from regime 1 ( \( r_{t} = 0 \)) to
regime 2 ( \( r_{t} = 1 \)) are presented as follows:
Assuming the parameters of bias are changing with unobserved regimes coded as $r_t$, the Markov switching model is defined:

$$price_t = \begin{cases} 
\alpha_1 price_{t-1} + \beta_1 bias_{t-1} + \gamma_1 stock_t + \lambda_1 sp_t + \delta_1 + \varepsilon_t \\
\alpha_2 price_{t-1} + \beta_2 bias_{t-1} + \gamma_2 stock_t + \lambda_2 sp_t + \delta_2 + \theta_t 
\end{cases}$$

Where $price_t$ is crude oil futures price at time $t$, $bias_{t-1}$ is the proxy of investor psychological bias, and $stock_t$ and $sp_t$ are control variables that include crude oil stocks and S&P 500 index which have been documented to influence oil prices. $\beta$ represents the effects of investor bias on oil prices. The error term is noted as $\varepsilon_t$. By estimating the Markov switching model, we can get the estimated coefficients shown in Table 2.

<table>
<thead>
<tr>
<th>Table 2. MSAR Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Coefficients</strong></td>
</tr>
<tr>
<td>$\delta$ (intercept)</td>
</tr>
<tr>
<td>$\alpha$ (previous oil price)</td>
</tr>
<tr>
<td>$\beta$ (investor bias)</td>
</tr>
<tr>
<td>$\gamma$ (oil inventory)</td>
</tr>
<tr>
<td>$\lambda$ (S&amp;P index)</td>
</tr>
<tr>
<td>Multiple R-squared</td>
</tr>
<tr>
<td>Residual standard error</td>
</tr>
</tbody>
</table>

**Model diagnostic**

- AIC: -605.31
- BIC: -501.30
- Loglike: 312.66

**Transition probabilities**

<table>
<thead>
<tr>
<th>Regime 1</th>
<th>Regime 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.76</td>
<td>0.30</td>
</tr>
<tr>
<td>0.24</td>
<td>0.70</td>
</tr>
</tbody>
</table>
Figure 1 shows the historical crude oil futures prices from January 2004 to February 2014. From this figure, crude oil prices pattern can be clearly visualized. The lower part of Figure 1 is the estimated smoothed probability using this empirical model. Table 2 shows the estimated coefficients of independent variables. Standard errors are reported in parentheses in Table 2. Because this paper focus on the influence of investor bias, we mainly analyze the change of $\beta$ in two regimes. Regime 1 represents the rising phase of crude oil price. Regime 2 means the decreasing phase of oil price. When the oil prices rise, one unit increase of investor bias can result in 0.11 changes of oil prices with controlling other independent variables unchanged. Quite to the contrary, when the oil price decreases, the response of oil price to investor psychological bias is significantly negative, and the coefficient is -0.21. From these estimated coefficients, the asymmetry between two regimes can be clearly observed. Transition probabilities are also reported in the lowest part of Figure 1. Using the formulas provided in Zhang and Feng 2011, I computed durations of rising and decreasing phases. The expected duration of rising and decreasing phases are 4.12 and 3.33 weeks, respectively. These results suggest that a rising phase lasts about 4 weeks, while a decreasing phase lasts about 3 weeks.

Furthermore, this paper analyzes the correlation among the proxy and trading volumes of four different types of traders in futures market to evaluate the new measure. The results show that this new proxy is strongly correlated with trading positions of small traders, with the coefficient being 0.5. Moreover, measures from the web are more timely in reflecting investors' psychological bias.

Discussion

This research leverages big data from the web to measure investor psychological bias in financial market. The measure has asymmetric influences on crude oil prices during the phases of rising and decreasing. This study is quite relevant with studies using big data to capture attention on specific stocks (Antweiler and Frank 2004, Da et al. 2011), and suggests that the asymmetric influences of big data to financial commodity prices. In this section, implications of behavior economics, contribution, and future work are discussed.

Implications of Behavior Economics

Since attention can be considered as a limited resource, a simple model proposed by (DellaVigna 2007) presents this scarce resource. Considering a commodity whose value $V$ is determined by the sum of two
components, a visible component \( v \) and an unobservable component \( o \). The formula is \( V = v + o \).

Owing to inattention, the consumer perceives the value to be:

\[
\hat{V} = v + (1 - \theta)o
\]  

(4)

Where \( \theta \) denotes the degree of inattention, with \( \theta = 0 \) as the standard case of full attention. The interpretation of \( \theta \) is that each individual sees the unobservable information \( o \). The inattention parameter \( \theta \) is a function of the saliences which belongs to \( 0 \) to \( 1 \).

Based on the psychology evidence, the inattention \( \theta \) is decreasing in the saliences \( s \) and increasing in the completing stimuli \( N \). Inattention is zero for a fully salient signal \( \theta(1, N) = 0 \).

This model suggests that three strategies to identify the inattention parameter \( \theta \). The first is to compute how the valuation \( \hat{V} \) responds to a change in \( o \). The derivative term can be computed to test for limited attention. The second is to examine the response of consumer valuation to an increase in the saliences. The third strategy is to vary the number of competing stimuli \( N \), and test whether it has an effect.

**Contribution**

This study contributes the new measurement of investor psychological bias leveraging big data crawled from the web search to the research community. The approaches and models may guide the future research on identifying the influences of investor bias on other commodities. For market practitioners including institutional and individual traders, this research presents a model to capture their trading attention, and suggests its relation with oil futures prices. Investor bias is an important influencing factor, and should be emphasized in prices analysis. In addition, web search data may become an important big data source to better understand investors’ trading psychology, and support their decision making.

**Future Work**

The primary future work is to process the data from the Internet forums, and use the same Markov switching model to investigate its influences on oil prices. For better depicting the bias, Eq. (4) will be further derived. Moreover, applying this method to examine the relationship between investor bias and other financial commodities should be worthwhile in the future work.

**Conclusion**

In conclusion, the investor psychological bias constructed using web search data series does influence financial commodity prices. The main contribution of this study is to construct new measure using web search data. The influences of this new measure on oil prices are asymmetric in two regimes of oil prices. The impact in phase of decreasing is stronger than that in rising phase. This paper discussed that this asymmetry may be brought by investor psychological bias like overconfidence and familiarity bias. This work will hopefully contribute to the recent literature on big data applications in the analysis of financial commodity markets.

**Acknowledgements**

The author thanks Prof. Ananth Srinivasan, Prof. Bernard Tan, Prof. Elena Karahanna, the associate editor, and the two anonymous reviewers for their constructive feedbacks, which enabled me to improve the paper significantly. The author thanks Professor Shouyang Wang and Professor Manel Baucells Alibes for their instruction. The author is also grateful to Professor Juan Feng for her insightful comments and suggestions.

**References**


