TRADING ON FORECASTED MACROECONOMIC INFORMATION

An Analysis of the S&P500's Overnight Reaction to Unemployment Rate Surprise and an Analysis of the Forecasting Power of Gallup's Unemployment Rate Estimates

---

BY

MIKHAIL ESIPOV

An honors thesis submitted in partial fulfillment

of the requirement for a degree of

Bachelor of Science

Undergraduate College

Leonard N. Stern School of Business

New York University

May 2010

Professor Marti G. Subrahmanyam   Professor Lasse H. Pedersen

Faculty Advisor   Thesis Advisor

Table of Contents

List of Figures

List of Tables

List of Appendices

1. ABSTRACT

The main purpose of this study is to examine the overnight impact of economic information released by the Bureau of Labor Statistics (BLS) on the S&P500 and then further study the possibility of trading on these releases through forecasted data collect by Gallup. Through a comparison of the actual BLS release with the pre-release market consensus, I discovered that there is negative correlation between the change in the unemployment rate and the change in the S&P500. After establishing a relationship between the unemployment *surprise* and the S&P500 return, I then attempt to forecast this change in the unemployment rate through a daily 1000 person survey conducted by Gallup. The Gallup data leads to two separate conclusions. First, although Gallup's full data set is consistently within 2% of the true release of the BLS, it is not sufficiently accurate in its current form for trading purposes. Second, when the data is adjusted to a 20 and older unemployment rate, an indicator that is also released by the BLS, the results conclude that Gallup's data can be used to forecast the BLS 20+ unemployment rate. The final portion of this thesis analysis survey derived macroeconomic statistical data on an emerging market basis to assess the potential for a long/short strategy based on the forecasted figures' deviation from the CIA World Fact Book.

2. INTRODUCTION

Through buying and selling goods and indulging in the latest services, consumers play an integral role in driving any economy towards growth or towards recession. Consumers are the ultimate drivers of growth in free cash flow through the expenses that they incur and consumers are the ultimate source of demand that thousands of financial analysts attempt to forecast in projecting dozens of firm metrics. But if the ultimate drivers of the economy are consumers, why does financial analysis not take into account more information pertaining to everyday consumers Similar to predicting the next U.S. President, when marketing researchers conduct exit polling, what is stopping a financial analyst from polling the consumers of the world in order to attain an informational advantage over the market?

One may argue that consumer data is already fully exploited by financial markets through anxiously awaited indicators such as the Consumer Confidence Index. Even election exit polls are priced into the market as analysts forecast the financial implications (fiscal policy, healthcare, etc) of each candidate's victory along with the likelihood of that victory. But I believe that more can be found. By asking the *right* amount of people the *right* questions, valuable information can be extracted from our everyday consumers.

For the purpose of this thesis, this data, for the most part, falls into one of two fundamental categories. Data can represent consumer opinion or outlook, which we nickname "crowd wisdom," or data can represent actual fact or state of being, which we nickname "crowd core." Although mainstream business-texts such as *Wisdom of the Crowds,* by James Surowiecki, suggest that given the right circumstances crowd wisdom can be the best wisdom, this is not a premise that most investors will be willing to risk their net worth on. Simply because a million consumers *think* Apple will beat market earnings expectations, does not necessarily

imply that Apple *will*, in fact, beat market earnings expectations. Crowd wisdom data, despite its qualitative convenience, is not sound financial information.

On the other hand, crowd core data represents actual fact: *actual* goods that have been purchased, *actual* services that have been indulged in, *actual* jobs that have been gained or lost. Crowd core data is the real-time data that will eventually be released by financial statements at the end of each company's quarter. On a day-to-day basis, consumers can tell us, first-hand, how many cars they purchased (non-durable goods), how much they spent today not including large purchases (durable goods), or where and when they intend to move or travel (real-estate and travel).

There are hundreds of firms that conduct frequent marketing research and collect this sort of data. Whether these firms are corporations studying the size of a particular target market or think-tanks hypothesizing about a new found socio-economic relationship, this data exists in abundance today. But in order to be exploitable financial/economic information, it must have three fundamental characteristics. First, it must answer the *right* questions - questions that lead to crowd core data. Second, this data must be privately collected in order to gain an informational advantage over the market. And third, this data must represent a *large* enough sample size in a statistically significant manner. Only if all three of these constraints are fulfilled does the data have potential to be financially useful for investment purposes.

Given these three constraints, the question now becomes, what specific data already exists in order to conduct a thorough study? For several years, Gallup, a well-known marketing research and management consulting firm, has conducted a daily 1000 person survey. The details of its methodology will be described in full later, but the important point is that this data asks many *right* questions and it is collected by a private firm. Whether or not the data collected

represents a large *enough* sample will be discussed later, but the take away is that this data is sufficient for a thorough study.

This study selects the U.S. unemployment rate as a proxy for information that can be forecasted by a private entity in order to attain a form of *insider information* on the state of the U.S. economy. Although there are many other important economic indicators, such as the consumer price index, consumer expenditure, or even payroll employment, the unemployment rate serves as an ideal candidate for this study. Not only does the information release happen afterhours which allows for a simple means of studying the impact of the new information, but Gallup has several months of historical estimates for this indicator. These factors allow me to not only study the overall impact of changes in the unemployment rate on financial markets, but also allow me to attempt to forecast the rate given Gallup's own daily poll.

3. HYPOTHESES

First Hypothesis

It goes without argument that the financial markets are ultimately a representation of the future outlook of the economy. Whether this statement is interpreted from a qualitative perspective or from the perspective of discounting future cash-flow generating power, financial markets are systematically dependent on the state and outlook of the economy. Integral to any functioning economy is the unemployment rate which in a way is a proxy for the state of macroeconomics relative to some long-run mean. One would expect that as the unemployment rate grows at a faster pace than what is forecasted by financial markets, then markets will interpret this positive *surprise* in a negative manner and experience a sudden sell-off. Similarly, if the Bureau of Labor Statistics (BLS) releases an unemployment rate below market expectations, a negative *surprise*, markets will interpret this as positive news and experience a sudden rally.

Given this basic macroeconomic intuition, I hypothesize that there exists a negative correlation between the surprise in the unemployment rate and the overnight return on the S&P500 on the monthly BLS release date. Although there is very likely to be noise on the monthly release date of the BLS, I believe that the unemployment rate *surprise* effects can be identified on a statistically significant basis. In addition, I expect that there exists a positive relationship between the unemployment rate *surprise* and the overnight return in 10 year treasuries due to the macroeconomic implications of a higher than expected unemployment rate. Furthermore, assuming that one is able to establish the BLS unemployment rate prior to the BLS release, I suspect that one can construct a profitable trading strategy which exploits this information asymmetry.

Second Hypothesis

Although the Gallup Daily Tracking poll has several key methodological differences when compared to the methodology of the BLS, I hypothesize that through using the optimal lag-time and the ideal weighted average of Gallup's daily poll, one can identify valuable financial information from this consumer data[1]. This information can be exploited by one of two methods. The first will study the statistical relationship between Gallup's unemployment rate and the return of the S&P 500. The second will attempt to forecast the monthly unemployment rate released by the BLS. Assuming the BLS unemployment rate release can be forecasted, one can implement the trading strategy from the first hypothesis to profit from the new economic information.

---

[1] Gallup's data is not currently available on a daily basis online, but the firm has kindly agreed to provide me with the necessary data to complete my analysis. Gallup has temporary hired me as a "Research Scholar."

4. UNEMPLOYMENT RATE METHODOLOGY OF THE BUREAU OF LABOR STATISTICS

The Bureau of Labor Statistics (BLS) of the U.S. Department of Labor has monitored the

U.S. unemployment rate since the later stages of the Great Depression. The sheer duration of

how long this economic indicator has been followed by the BLS is an indication of its

importance to the state of the economy. Jobs are an essential component to any functioning

economy. Unemployed individuals cannot collect a wage and thus cannot consume goods as

readily as employed individuals can. Furthermore, when the purchasing power of these workers

is lost, it can lead to unemployment for yet even more other workers.

As defined by the BLS, there are three fundamental categories that an individual in the

U.S. can fall into. If the person has a job, whether a full-time or part-time job, this person is

considered employed. If a person is jobless, looking for a job, and is currently available for work,

then this person is considered unemployed. If the person is neither employed nor unemployed,

then the person is not considered to be in the labor force[2].

Every month, the BLS conducts a sample survey called the Current Population Survey

which measures the extent of unemployment in the country. The BLS contacts 60,000

households which translates into approximately 110,000 individuals. The 60,000 households are

selected in such a manner to be representative of the entire U.S. population. Every month, one-

fourth of the households in the sample are changed, such that no household is interviewed more

than four consecutive months. After a household is interviewed for four consecutive months, it is

removed from the sample for eight months, and is then again interviewed for the same four

---

[2] The following link provides a more thorough explanation of who is considered employed vs. unemployed vs. not in the labor force: http://www.bls.gov/cps/cps_htgm.htm#why

calendar months a year later. After this second four months of interviewing, the sample is removed from the sample for good. According to the BLS, "this procedure results in approximately 75% of the sample remaining the same from month to month and 50% from year to year."

For each month, the BLS selects a target week which according to its website is "usually the week that includes the 12th of the month." The data collected during this reference week is then processed for several weeks until it is released to the public on the first Friday of the following month (unless that Friday falls on the first of the month in which case it is released the following Friday).

In addition to the standard unemployment rate, the BLS also announces an unemployment rate for people 20 years of age and older, total number of employed, the distribution across regions and industries, revisions to past unemployment announcements, employment totals, weekly and hourly earnings, and weekly hours worked. This thesis focuses solely on the standard unemployment rate and the 20 years of age and older releases.

5. HOW DOES THE BLS SURPRISE IMPACT FINANCIAL MARKETS?

Now that the reader has established an understanding of the methodology and release schedule of the BLS, it is possible to study the relationship between the unemployment rate and the financial markets. I have chosen to focus this analysis on the S&P500 for two primary reasons. First, the S&P500 is composed of 500 large-cap companies that are nationally diversified and the index is an excellent proxy for the overall state of the economy. Second, the S&P500 is a highly liquid index where an overnight change can not only be easily monitored,

but can logically be connected to the BLS information release due to the information release timing.

Before studying the impact of the BLS release on the overnight return in the S&P500, it is important to understand the "big-picture" relationship between the unemployment rate and the S&P500. From the Figure 1 below, one immediately observes a negative relationship between the unemployment rate and the S&P500 in line with my first hypothesis. It is important to mention that the unemployment rate is a *level* indicator that fluctuates overtime around some long-term average whereas the S&P500 is a *growing* indicator with a *positive* drift[3]. Thus, although for this selected 10 year sample the S&P 500 appears to be relatively flat, for a longer sample it should have a positive drift that may not display the same negative correlation with a level unemployment indicator.

Figure 1: U.S. Unemployment Rate vs. S&P500
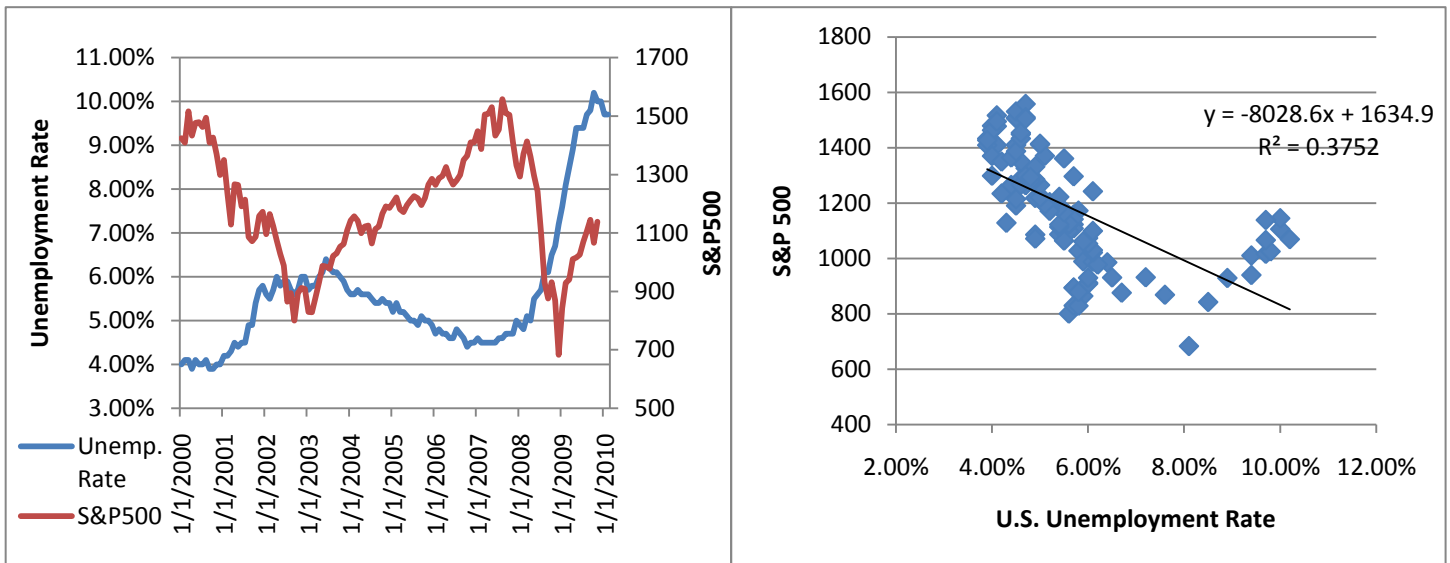
A simple regression using 10 years of historical data indicates the following relationship has a T-statistic of -8.38 and a P-value of 0.0. Clearly, there exists a strong negative relationship.

---

[3] Some may argue that the unemployment rate will experience its own positive drift as the U.S. economy systematically changes.

Also, please note that the unemployment rate is quoted as a percentage. A 1% increase in the unemployment rate should result in an 80 point drop in the S&P500.

Next, it is important to examine the historical performance of S&P500 on an overnight basis. The overnight return is calculated as follows:

———————

From the Figure 2 below it becomes evident that overnight return in the S&P500 is most volatile during periods of economic uncertainty (i.e. 2000-2003 and late 2007-2010).

Figure 2: Historical Overnight % Change in the S&P500
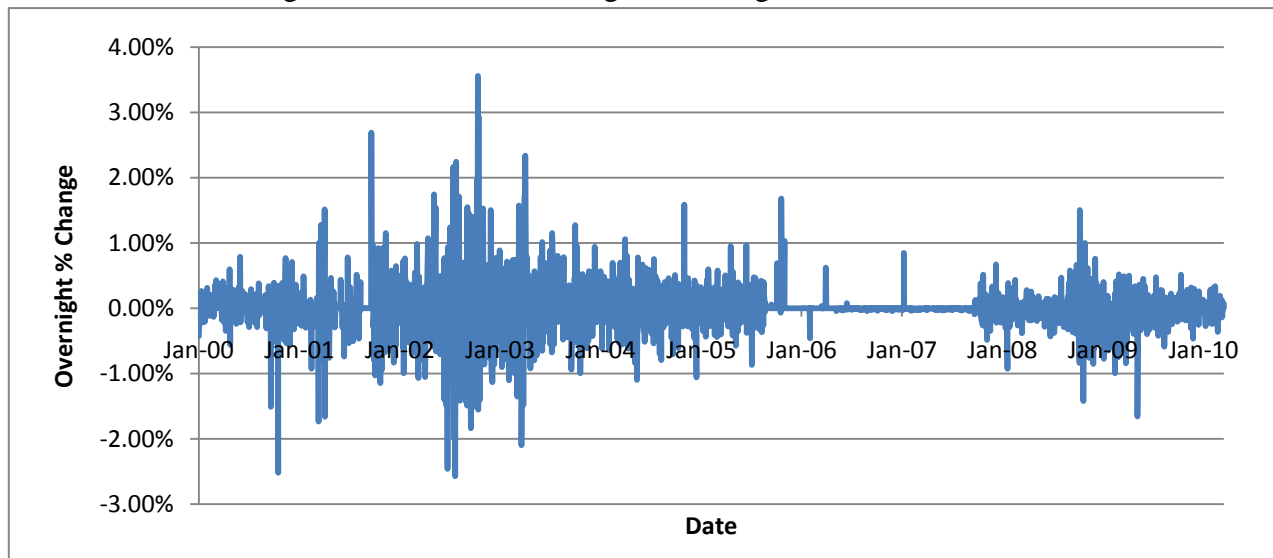
Although the opening price of the S&P500 may seem like a simple dataset to find, there are many flawed datasets which do not differentiate the difference between the open and close prior to 2008 (i.e. Bloomberg and Yahoo Finance). This data was collected from DataStream and although there are periods of zero overnight movement, as evident in Figure 2 from January 2006-January 2008, DataStream is the most accurate information I could obtain.

As explained in the BLS methodology section, the BLS releases its unemployment rate on the first Friday of every month unless this Friday falls on the first of the month in which case the data is released on the 8th of the month. In order to establish a proxy for the unemployment rate *surprise* I have subtracted the average consensus collected from Bloomberg's survey of economists from the actual rate released by the BLS. A positive surprise indicates that the unemployment rate turned out to be larger than expected and a negative surprise indicates that the unemployment rate turned out to be smaller than expected. I collected 10 years of this surprise data from Bloomberg which translates into 118 data points for regressions.

Although there have been studies on the relationship between the unemployment rate and financial markets, their conclusions vary. Boyd, Hu, and Jagannathan (2005) prove that "an announcement of rising unemployment is good news for stocks during economic expansions and bad news during economic contractions." Although Boyd et al. briefly examine an unemployment rate surprise relative to some market consensus, the paper's statistical model and regressions do not incorporate any notion of this *surprise* relative to a market consensus[4]. This implies that although rising unemployment may be interpreted as good news during a boom, it is rising relative to the previous month's unemployment rate and not relative to market consensus.

My initial regression takes into account the full set of data points which resulted in a T-statistic of -1.15, a P-value of 0.251 and the following relationship:

---

[4] The unemployment rate surprise is "based on consensus forecasts published in the *Wall Street Journal* from February 1992 to August 1994 in a previous study by Fleming and Remolona (1998).

Although this equation above displays the hypothesized negative relationship, the statistical significance of this regression is quite weak[5]. Figure 3 below displays this same relationship graphically:

Figure 3: Overnight %Change in S&P on BLS Release Date vs. %Unemployment Rate Surprise
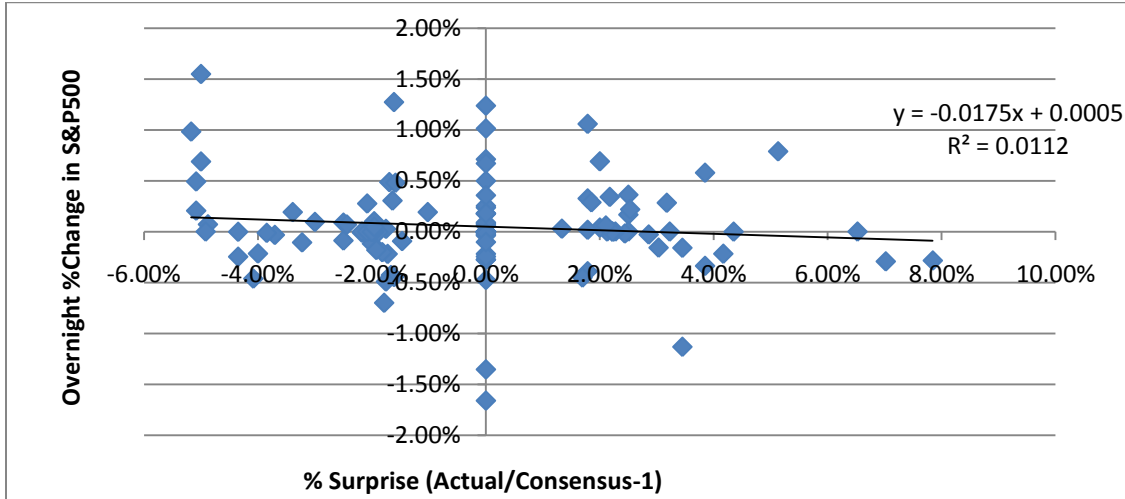
Figure 4 below displays a fundamental problem with the data that is due to the market's overall accuracy in predicting the BLS's unemployment release. Of the 118 data points, 45 of them correspond to a surprise of 0%.

Figure 4: Unemployment Rate Surprise Variations

---
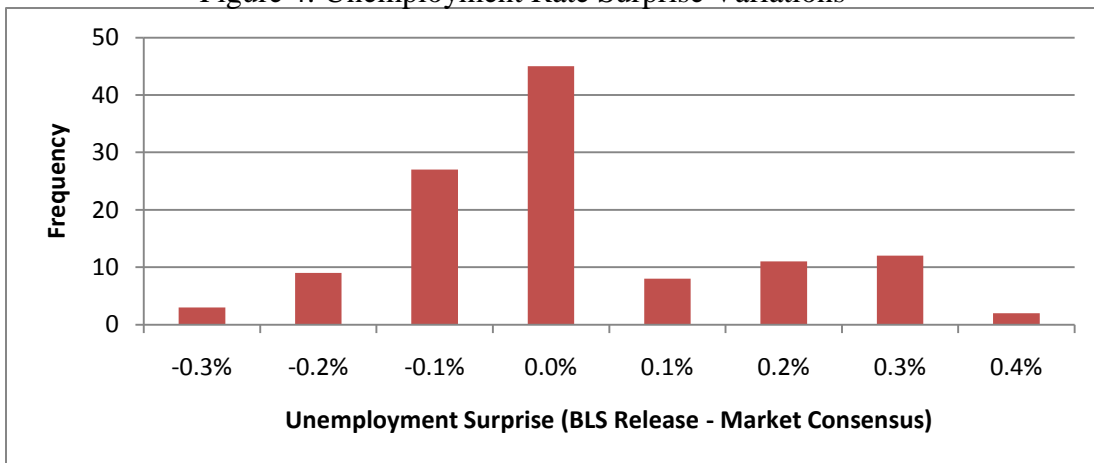
[5] For the sake of complete analysis, I also analyzed the day-to-day performance of the S&P500 on the date of the BLS release. By comparing the close at time to the close at time t-1, I discovered that there was far too much noise that completely overpowered the impact of the unemployment surprise. Furthermore, the regression showed a positive relationship (=0.00097+.0329*surprise)*, rather than a negative one, and a P-value of 0.466.

Since one would not expect any significant market movement in accordance with a surprise of 0%, I decided to remove these data points from the set. The table below displays the results of the same regressions that now omits certain levels of surprise and each regression's corresponding P-values:

Table 1: Regressions Output of Unemployment Surprise vs. S&P Overnight Return

Selected Dataset	Constant	Unemployment Coefficient	P-Value	Number of Data points
Full Dataset	0.000486	-0.398	0.144	118
Dataset without 0	0.000572	-0.392	0.119	73
Dataset without 0, +/-.1%	0.000855	-0.612	0.053	38
Dataset without 0, +/-.1%, +/-.2%	**0.00223**	**-1.17**	**0.025**	**12**

As evident from the table above, as one narrows the dataset into larger and larger surprises, the unemployment coefficient becomes more pronounced and the statistical significance of the regression increases substantially. For example, if one narrows the dataset to deviations of +/-.3% and greater, the regression holds for greater than 95% confidence. Figure 5 and the regression below display these results in more detail.

Figure 5: Extreme Unemployment Surprises vs. the Overnight Return of the S&P500
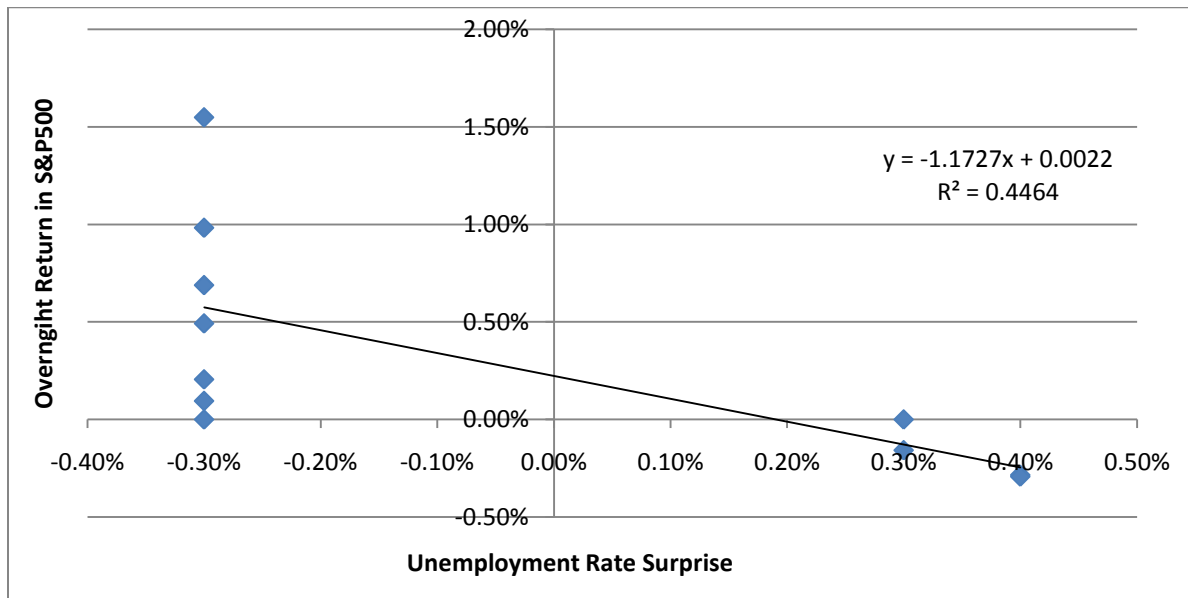
Based on this regression, if the BLS releases an unemployment rate of .3% greater than the market consensus, the S&P will contract by approximately .35% with over 95% confidence. Furthermore, this regression indicates that on average the unemployment surprise explains approximately 45% of the overnight movement in the S&P500 on the day of the BLS release.

Bond Market Analysis: Impact on the 1 Year Treasury Bill and the 10 Year Treasury Notes

For the sake of a more complete study of the effects of the unemployment rate, I have repeated parts of the above analysis on the U.S. bond market. I have chosen to study the day-to-day returns of the 1 year Treasury bill and the overnight return of the 10 year Treasury note. Similar to like the S&P500, the 1 year Treasury bill and 10 year Treasury note are highly liquid and are ideal candidates to monitor the effects of the unemployment surprise.

I have studied the 1 year Treasury bill on a day-to-day basis as posted by the Federal Reserve Bank of St. Louis. Similar to the regression conducted on the full S&P500 dataset, the regression of the complete 1 year Treasury bill dataset leads to inconclusive results. Once again, the high frequency of 0 surprises is likely the cause of these results. Thus, I have broken up the dataset by the size of the surprise and the regression outputs as summarized in Table 2 below.

Table 2: Regressions Output of Unemployment Surprise vs. 1 Year Treasury Overnight Return

Selected Dataset	Constant	Unemployment Coefficient	P-Value	Number of Data points
Full Dataset	0.000087	0.05921	0.148	114
Dataset without 0	-0.000239	0.3841	0.197	78
Dataset without 0, +/-.1%	**-0.0000712**	**1.4368**	**0.07**	**30**
Dataset without 0, +/-.1%, +/-.2%	-0.0004857	2.101	0.189	10

The first observation is that unlike for equity markets, a positive unemployment surprise leads to a positive return in the 1 year Treasury. In this case, the regression results improved after dropping both the 0 and +/-.1% surprises, but not after dropping the +/-.2% surprises. Although a p-value of .07 is not considered statistically significant, it shows that there exists a strong positive relationship between the unemployment surprise and the day-to-day return of the 1 year Treasury bill.

Now that we have analyzed the relationship between the 1 year Treasury bill and the unemployment rate surprise, we will analyze how this surprise affects the 10 year Treasury note. As evident from Figure 6 below, the 10 Year Treasury note, just like the S&P500, experiences heightened volatility during periods of higher economic uncertainty (i.e. 2000-2003 and late 2007-2010).

Figure 6: Overnight Return of the 10 Year Treasury Note
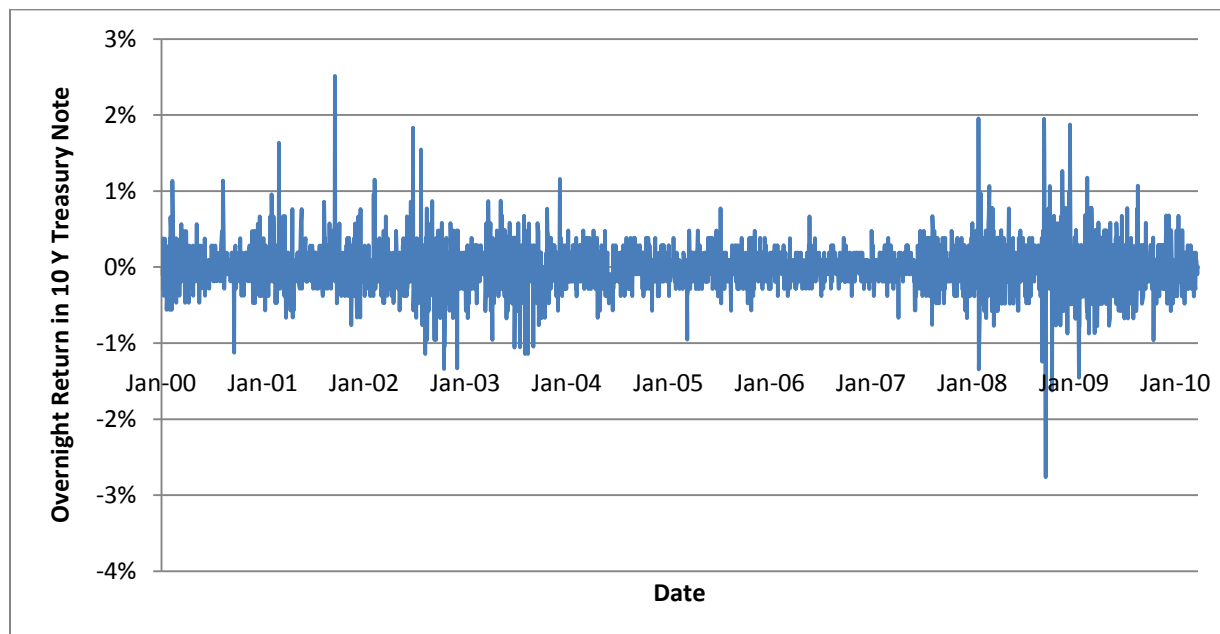
Once again, a regression of the full dataset leads to inconclusive results. Table 3 below summarizes the regression results for datasets depending on which unemployment rate surprises were excluded.

Table 3:Regression Outputs of the Unemployment Rate Surprise vs. 10 Year Treasury Note Overnight Return

Selected Dataset	Constant	Unemployment Coefficient	P-Value	Number of Data points
Full Dataset	0.000083	-0.01	0.954	118
Dataset without 0	0.000203	-0.003	0.988	73
Dataset without 0, +/-.1%	0.00013	0.126	0.515	38
Dataset without 0, +/-.1%, +/-.2%	**0.000195**	**0.291**	**0.257**	**18**

Although the full dataset shows a negative relationship as with the S&P500, this negative relationship disappears after removing deviations of 0% and +/-.1% from market consensus. When focusing solely on deviation of .3% or more, the resulting regression becomes:

Thus, an unemployment surprise of +0.3% would result in an expected overnight return of .1% with approximately 75% confidence. This implies that with a positive surprise the bond price tends to increase overnight which implies that the 10 year T-Note yield tends to contract. The regression confirms the initial hypothesis of a positive relationship between the unemployment rate surprise and the overnight return on the 10 year T-Note.

Although I have not studied the exact causality of the relationship, it is potentially due to market expectation that higher unemployment leads to a monetary policy of lower rates. Lower rates in turn lead to increased prices which may explain the positive relationship. Additionally,

this relationship may be explained due to the common negative correlation between equity

markets and debt markets. As equity markets begin to sell-off as a result of the negative news of

a positive unemployment rate *surprise*, capital will tend to flow into debt markets as a sort of

"flight-to-quality." Similarly, capital will tend to flow from the debt markets to the equity

markets in the result of the positive news of a negative unemployment rate surprise due to better

than expected economic conditions.

Trading on the Unemployment Rate Surprise

After a thorough analysis of how the unemployment rate surprise affects both the equity

markets and the debt markets, let us now *assume* that we have a way of determining what the

BLS will release prior to each month's release date. Whether I am an economist that is very

confident in what the BLS will release or whether I have a means of forecasting what the rate

will be through some form of marketing research (second hypothesis), let us examine the results

of several trading strategies that exploit this hypothetical information asymmetry.

First of all, it is important to note that since the unemployment rate surprise has a

stronger relationship with the equity markets than with the debt markets, equity will be the

market of choice. Second, the following analyses will use the same 10 years of historical

surprise data as in the regressions. Third, due to the increasing statistical significance associated

with larger unemployment rate surprises, trading strategies should use higher leverage for higher

surprises. Over the past 10 years, the unemployment rate surprise falls into one of eight

categories: +/-.3%, +/-.2%, +/-.1%, 0% and +.4%[6]. By manipulating the amount of leverage

---

[6] There has never been a -.4% surprise in the unemployment rate.

taken for higher deviations from expectation, a trader can exploit the added confidence of market reaction associated with higher deviations.

Based on the negative relationship between the unemployment rate surprise and the overnight S&P500 return established in the first part of this section, this analysis will begin by focusing on the results of the following strategy:

- If the surprise is negative, go long the S&P500

- If the surprise is positive, go short the S&P500

- If the market is correct and there is a surprise of zero, do not invest.

This strategy, if implemented once a month, assuming no investment or cash interest on days other than release days, yields an annualized return of .2%. Clearly for any fund to begin considering this strategy, leverage is a necessary component.

By breaking down the dataset by the size of the surprise we arrive at the following results in Table 4:

Table 4:Statistical Characteristics of Unemployment Rate Surprise

Surprise Size	Average Return	Standard Deviation	Sharpe Ratio
0.00%	0.00%	0.00%	NA
0.10%	-0.04%	0.35%	-0.13
0.20%	-0.04%	0.37%	-0.10
0.3%+	0.43%	0.48%	0.91

Based on this analysis, it appears that the only viable strategy is to focus on the large unemployment rate surprises. In fact, if a trader were to trade on the smaller deviations (+/-.1% and +/-.2%), he or she would lose money on average. Additionally, trading on large deviations is a better strategy based on the statistical significance previously determined in the regressions.

Table 5 focuses only on the large unemployment rate surprises and the return of an unlevered position.

Table 5:Overnight Returns of Large Unemployment Rate Surprises

Release Date	Difference	Overnight S&P500 Return	Overnight Large Surprise Return	Annualized Return[7]
9/7/2001	0.30%	0.00%	0.00%	0.00%
2/8/2002	-0.30%	0.21%	0.21%	52.92%
3/8/2002	-0.30%	0.98%	0.98%	246.96%
6/7/2002	-0.30%	0.00%	0.00%	0.00%
9/6/2002	-0.30%	1.55%	1.55%	390.60%
10/4/2002	-0.30%	0.49%	0.49%	123.48%
2/7/2003	-0.30%	0.69%	0.69%	173.88%
6/6/2008	0.40%	-0.28%	0.28%	71.57%
9/5/2008	0.40%	-0.29%	0.29%	73.84%
11/6/2009	0.30%	-0.16%	0.16%	39.82%
2/5/2010	-0.30%	0.10%	0.10%	23.94%

From Table 5, the first conclusion is that a strategy that trades only on large deviations has not lost money in the past 10 years; the market's response to a significant unemployment rate surprise has *always* moved in accordance with the first hypothesis. Past performance is no indication of future performance, but nonetheless, this consistent trend and the statistically significant regression justify high leverage rates for these overnight trades. The zero overnight return on 9/7/2001 and 6/7/2002 are likely to be the result of data error from Datastream. If these two data points are dropped, the average increases to .53% and the Sharpe Ratio increases to 1.11.

Although the annualized returns may appear to be great, the problem with this strategy is that it can be implemented only very infrequently. The unemployment surprise is only large enough for *confident* trading during times of rapid economic change. The first seven large

---

[7] Annualized return is calculated as the overnight return multiplied by 252.

differences occurred in 2001-2003 and the remaining four occurred in 2008-2010. Thus from 2003-2008, for over 5 years, this strategy would have remained entirely still. Clearly this strategy cannot be implemented as a sole investment approach, but rather as an addition to some event-based investment approach. The overnight annualized returns from large unemployment surprises are clearly large and with the right amount of leverage can be fantastic trades. The table and charts in Appendix A summarize a variety of simple trading strategies that vary based on the amount of leverage placed on a particular BLS unemployment release.

6. Extracting Financial Information from Gallup's Daily Tracking Survey

This section is focused on testing the second hypothesis to determine whether or not Gallup's Daily Tracking survey provides new and exploitable financial information to the markets. In order to test this hypothesis, one must first attain a better understanding of the difference in Gallup's methodology versus the methodology of the BLS. After this is established there are two fundamental approaches to understanding how Gallup's findings can provide valuable information to the financial markets. The first approach is to study how movement in Gallup's day-to-day unemployment rate, through a variety of moving averages and lags, correlates to movement in the S&P500. The second approach is to analyze how accurate Gallup's unemployment rate is to the rate released by the BLS. Both of these strategies have simple trading strategies that can be implemented to take advantage of the information asymmetry.

An Analysis of Gallup's Daily Tracking Methodology

Gallup's Daily Tracking survey is a daily poll of a sample of 1000 people. The length of the poll has varied over the years, but on average it answers a total of 100 questions varieties in approximately 15 minutes. Gallup has polled Americans for decades, but the Daily Tracking survey only started in 2008. Gallup began testing a new set of employment questions in January 2009 and in April 2009 it launched the employment series. In November 2009, Gallup decided to make a few tweaks to the initial questions and did split sample testing in November and

December to optimize its methodology. In December 2009, Gallup adopted the latest and greatest series of questions which did not change the theoretical basis of its poll, merely several wording changes. This study uses a combination of the two methodologies in order to increase sample size. At the time of writing this thesis, using only the new Gallup methodology would have only allowed for three months of comparison. With time, as Gallup's new methodology collects more data points, the study should be repeated as the statistical significance of these findings will likely change.

Gallup's definition of the unemployment rate follows the same format explained in Section 4 as outlined by the BLS: first measure the size of the workforce and then divide the total number of unemployed by that size of the workforce. Gallup's definition of unemployed vs. employed also follows the same standards as the BLS. The samples are stratified by region and then weighted to represent the entire U.S. population before calculating a final figure. Finally, Gallup's policy has been to only poll people that are older than 18 years old.

Although the sheer sample size of the BLS's methodology likely implies a more accurate unemployment estimate than Gallup's, the key problem is the delay with which it reports. In today's drastically changing economic environment, a three week delay in the unemployment rate can represent a dramatic difference relative to what *true* unemployment is. That being said, Gallup's daily figures could represent a more accurate approximation of what the *true* unemployment rate is if one assumes that its sample size is large enough. It is important to note that even if Gallup's methodology is a better approximation, in all likelihood, the market currently prices in the BLS figures and thus an understanding of true unemployment may be useless if the market never accepts it as true. Nonetheless, if despite the methodological difference Gallup's figures can forecast the BLS's expected release, then a clear trading strategy can be implemented.

Table 6 below summarizes the key differences between the methodologies of Gallup and

the BLS:

Table 6:Unemployment Rate Methodology Comparison of Gallup versus BLS

	Gallup	**BLS**
Sample Size	1000 people night, 30,000 people per month	60,000 households for one week of every month
Respondent Type	Gallup asks individuals about their own employment situations, not those of the entire household	The BLS communicates generally with one individual about the employment situation of all members of a household
Frequency	Daily	Monthly
Data Release Date	An ongoing rolling average which is released daily through its website as a 30-day rolling average	Data released once a month which represents approximately a 3 week delay in what true unemployment is
Historicals	Less than a year of data, with only 3-4 months the latest version	BLS has collected the unemployment rate since the later stages of the Great Depression
Age	Gallup only polls 18+	BLS standard unemployment figures are 16+
Margin of Error[8]	+/-3% for daily, +/-.7% for the 30-day average	+/-.2%

In order to gain a better understanding of Gallup's daily data, I have conducted a time

series analysis using an Autoregressive Moving Average (ARMA) model. Although several

variations of the ARMA model yielded statistically significant results, the strongest results came

from an ARMA (1, 1): one autoregressive term and one moving average term. The model

follows the following format and the results are displayed in Table 7:

---

[8] Margin of Error was taken as given from a Gallup Representative.

Table 7:Autoregressive Moving Average Model Outputs

	Coefficient	T- Statistic	P- Value
Constant	0.00378 (α)	22.08	0.00%
Autoregressive	0.9585 (β)	49.97	0.00%
Moving Average	0.5184 ( )	9.22	0.00%

Based on these outputs, several important points become clear. First, all of the T-statistics show statistical significance and that the coefficients are in fact non-zero. Second, the autoregressive coefficient of nearly 1, without the moving average term, implies that changes in the Gallup survey follow a random walk; they are almost unpredictable. However, the moving average term suggest that there is predictability in the movement. If Gallup's unemployment rate jumped up substantially today, one can expect this jump to be reversed by approximately 50% the following day.

Gallup Unemployment Rate versus Financial Markets

When one looks at Gallup's daily unemployment rate, the immediate observation is that the data is quite volatile on a day-to-day basis. The following chart shows the daily data along with the day-to-day log change in Gallup's unemployment data[9].

---

[9] The chart indicates extremely high volatility in the unemployment rate throughout February. When asked about the reason for this fluctuation, Gallup indicated it is most likely the result of the very heavy snowfall that hit the East coast. "This could result in quite a bit of fluctuation in employment numbers – either because of people taking work (shoveling snow, etc) or because of people not being able to get to work."

Figure 7: Gallup's Daily Unemployment Rate and Daily Log Change
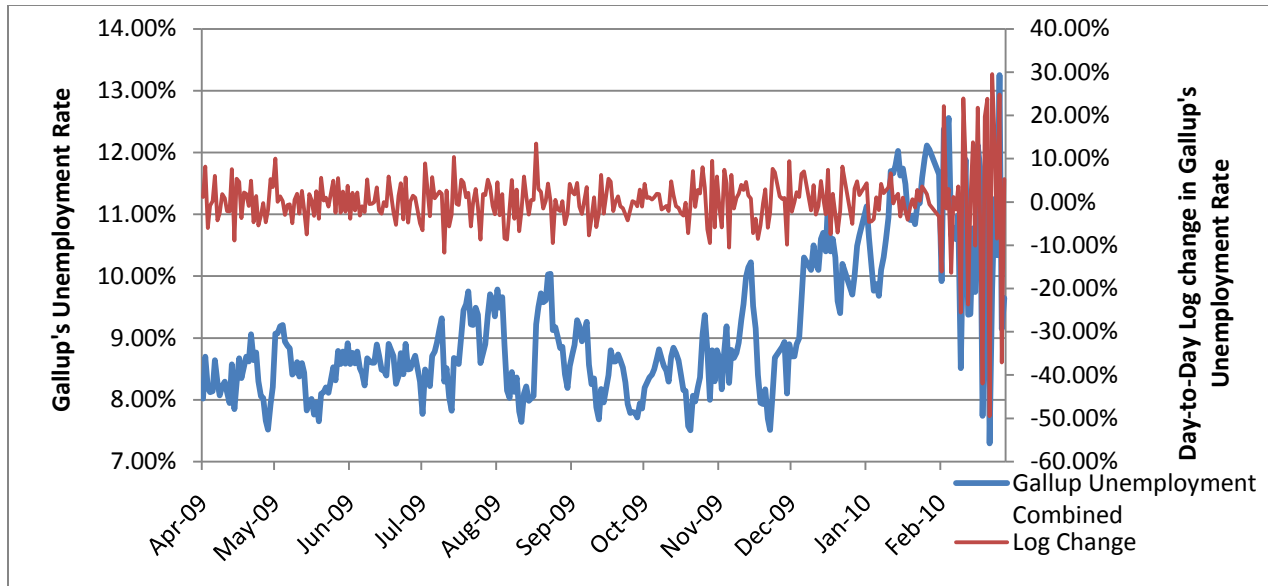
Based on this data, the daily standard deviation calculated with 11 months of Gallup's unemployment rate data is 1.13% and has a long term average is 9.08%. Gallup's unemployment rate standard error can be calculated using the following formula:

$$\overline{\overline{}}$$

This binomial standard error formula is appropriate if we assume that the unemployment rate has one of two states. Either a data point is part of the unemployment rate, p, or it is part of the employment rate, 1-p. Given that a sample size of 1000 represents people unemployed, employed and not in the labor force, it is important to change N to be the number of people, out of 1000, that are in the labor force. Gallup's surveys, on average indicate that approximately 65% of the U.S. population is considered part of the workforce and therefore N is set as 650. This implies that the binomial standard error for the full dataset is 1.13%. This value coincides with the standard deviation of Gallup's unemployment rate which implies there is minimal sampling error in Gallup's methodology.

Although the data contains an appropriate amount of sampling error, the use of a weighted average is necessary to smooth the large deviations in unemployment rate. Without the use of a weighted average a sudden 2-3% movement in Gallup's unemployment rate could lead a trader down a potentially dangerous path. One should expect that longer moving averages will have higher the statistical significance up to a certain level. Furthermore, it is highly likely that a lagging moving average may prove to be most statistically significant due to lag with which the BLS releases the unemployment rate after it is measured.

The approach I have chosen to forecast the S&P500 returns with Gallup's unemployment rate data is based on the following relationship:

This equation takes the percentage change in the Y week moving average from time T-Y to time T as a measure of the change in the unemployment rate. A Y week delay between the moving averages is appropriate because it prevents data points from being used in both moving averages. Based on the linear relationship above, this unemployment rate change is then compared to the change in the S&P500 over the next X weeks. For example, if Y=3 and X=2, this implies a the three week percentage change in the three week unemployment rate moving average compared to the return in the S&P500 over the next two weeks. Figure 8 below displays this setup visually.

Figure 8: Illustration of S&P500 Return Forecasting Methodology
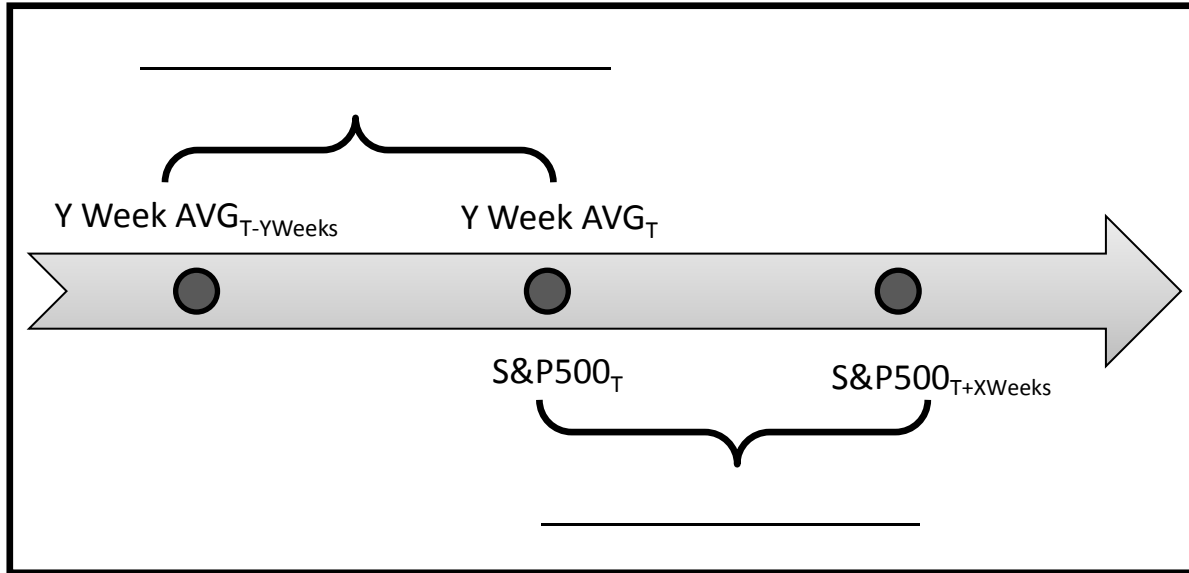
Tables 8 and 9 below summarize the 25 regression results of five unemployment rate moving averages (Y) with the return in the S&P500 over 5 separate intervals (X).

Table 8: S&P500 Return Regression Output – P-Values

| | | S&P500 Return Period | | | | |
		Next Day	1 Week	2 Week	3 Week	4 Week
Moving Average	1 Week	0.579	0.317	0.842	0.55	0.637
	2 Week	0.17	0.76	0.156	0.365	0.686
	3 Week	0.474	0.2	0.062	0.263	**0.034**
	4 Week	0.968	0.983	0.225	**0.003**	**0.003**
	EWMA	0.453	0.962	0.689	0.245	0.37

Table 9: S&P500 Return Regression Output – Coefficients

| | | S&P500 Return Period | | | | |
		Next Day	1 Week	2 Week	3 Week	4 Week
Moving Average	1 Week	0.0069	0.0991	-0.027	-0.094	-0.091
	2 Week	0.0336	0.071	-0.433	-0.308	0.16
	3 Week	0.0246	-0.422	-0.822	-0.549	**-1.22**
	4 Week	-0.0017	-0.009	-0.702	**-2.03**	**-2.38**
	EWMA	0.0335	-0.017	-0.18	-0.614	-0.565

The results of these 25 regressions lead to several key conclusions. First, an S&P500 return period of at least 2 weeks is necessary in order to observe the expected negative correlation between the unemployment change and the S&P500 return. Second, with a return period of two to four weeks, the statistical significance of the regressions tends to increase with longer moving averages. For example, a four week moving average with a three or four week S&P500 return period results in a p-value of .003, clearly statistically significant[10]. The regression coefficients for these two regressions are -2.03 and -2.38 accordingly. This indicates that if Gallup's 4 week unemployment rate moving average were to decline by 1%, one should expect at least a 2% increase in the S&P500 three or four weeks later with statistical likelihood[11]. Furthermore, a return period of three or four weeks is appropriate given the three week processing time that the BLS takes to release the unemployment rate found from its target week. The processing time acts essentially as a lag between the actual unemployment rate and the time it is absorbed by the market.

Five of the 25 regressions incorporated an exponentially weighted moving average. Although these results follow a similar pattern to the straight moving averages (decreasing p-value and negative coefficients for three or four weeks of lag), the p-values are not low enough for statistical significance. The exponential moving average was determined by fluctuating λ in the following equation in order to maximize the R-squared:

---

[10] It is important to note that the above findings may likely be significantly altered due to the positive serial correlation in the residuals. While writing this thesis I did not have access to any software that could adjust for standard error with a Newey-West estimator. Thus the above results should be interpreted with caution.

[11] 4WS&P500 Change=.0328-2.38 4W Gallup Unemployment Rate with 4 week Lag

Based on the statistical significance, the results of these regressions indicate that Gallup's data does have the ability to forecast the change in the S&P500. The following chart plots out the S&P 500 along the left axis and Gallup's 4W moving average along the right axis.

Figure 9: Gallup's 4 Week Unemployment Rate Moving Average vs. S&P500
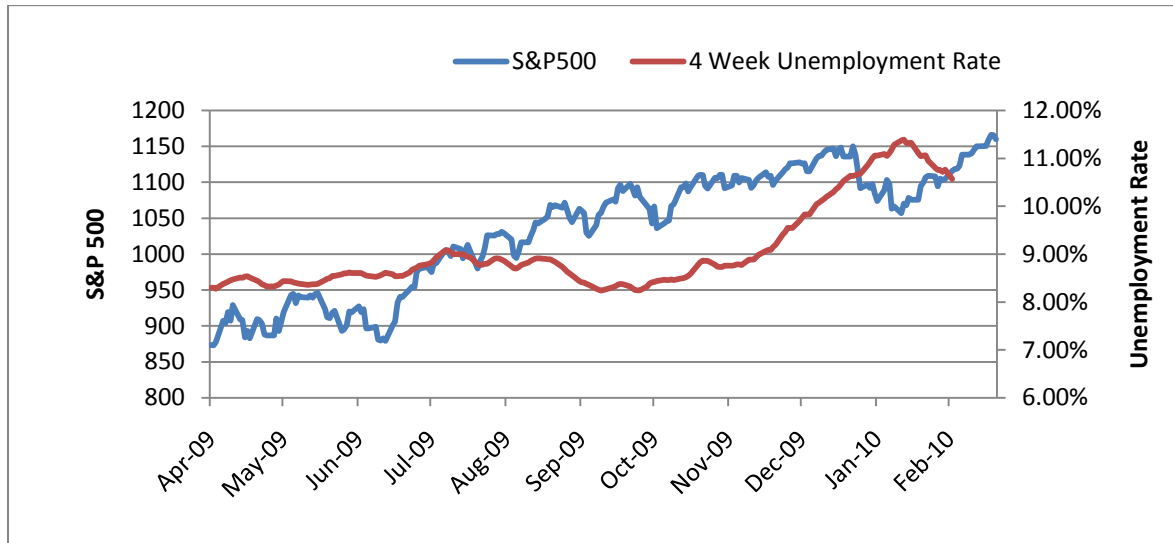
There are two key points to focus on. First, in December of 2009, Gallup's 4 week moving average climbs from 9% to 10%. Accordingly, over the next 4 weeks the S&P500 declines from 1150 to 1075. The second date of interest is February of 2010. When Gallup's 4 week unemployment rate declines from 11.5% to 10.5%, the S&P500 increases from 1100 to 1175 over the next 4 weeks.

These results prove my hypothesis that consumer marketing research data can provide valuable information to financial markets. Although Gallup's unemployment rate is a relatively short time series, the fluctuations in Gallup's 4 week unemployment rate moving average have a statistically significant relationship to the return in the S&P500 over the following 3 to 4 week period.

Gallup Unemployment Rate versus the BLS Unemployment Rate

The second alternative to exploit the informational advantage of the daily Gallup unemployment data is to use it to forecast what the BLS will release on the first Friday of every month. If Gallup's data can be used to forecast the BLS's release, then a clear trading strategy can be implemented. This strategy would first forecast the unemployment rate surprise and then would take a market position in accordance with one of the trading strategies discussed in the results from the first hypothesis of this thesis.

Based on the BLS methodology of picking a target week, analyzing the data, and releasing the unemployment rate to the public the beginning of the following month, there are four different trading methodologies. The first calculates Gallup's own seven day target week average based on what the BLS target week is. Although this method studies the same exact week that the BLS focuses on, it is most likely inadequate since a 7,000 person sample is not sufficient to forecast a 60,000 household sample (about 150,000 people). In order to attempt to compensate for this insufficiency, the second methodology averages one week prior and one week after the "target week" for a total of 21,000 people. The third methodology takes this adjustment one step further by incorporating two weeks prior to the target week and two weeks after the target week. Although this third approach now has a total of 35,000 data points, it may be inaccurate at forecasting the BLS's figure due to the natural unemployment fluctuations in the two weeks prior and after that will alter the final results. Too long of an average will increase sample size at the expense of capturing outdated data. The fourth strategy simply uses the exponentially weighted moving average up to the end of the target week. The mock calendar below summarizes the first three forecasting methods. The grey days indicate the BLS target week and the blue diagonal and bolded date indicates the data release date of the prior month.

	Sun	Mon	Tues	Wed	Thur	Fri	Sat
		1	2	3	4	5	6
Method 1, 1 Week	7	8	9	10	11	12	13
	14	15	16	17	18	19	20
	21	22	23	24	25	26	27
	28	29	30	1	2	3	4
Method 2, 3 Weeks	5	6	7	8	9	10	11
	12	13	14	15	16	17	18
	19	20	21	22	23	24	25
	26	27	28	29	30	1	2
	3	4	5	6	7	8	9
	10	11	12	13	14	15	16
	17	18	19	20	21	22	23
Method 3, 5 Weeks	24	25	26	27	28	29	30
	1	2	3	4	5	6	7
	8	9	10	11	12	13	14
	15	16	17	18	19	20	21
	22	23	24	25	26	27	28
	29	30	1	2	3	4	5

Before conducting the actual regressions, it is important to address a few limitations. First, although the four methods do their best to overcome the methodological difference between Gallup and the BLS, they remain estimates. Second, Gallup has only been collecting this data since April 2009 which translates into 11 months of comparison or 10 month-over-month changes for comparison. Due to such a small data sample and particularly the large amount of fluctuation in the unemployment rate in 2009-2010, this data may not be representative of the *true* correlation between BLS's and Gallup's unemployment rate. Figure 10 displays the BLS's actual unemployment release along with the estimates of all four methodologies.

Figure 10: BLS Standard Unemployment Rate vs. Gallup Averages
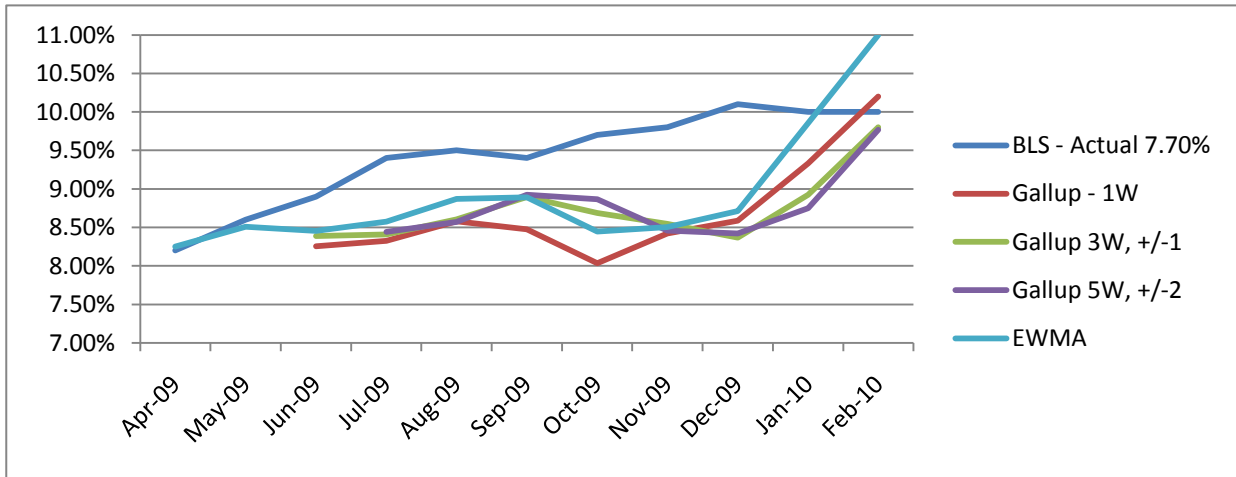
By looking at percentage change month-over-month, the four regressions have the

following format and lead to the regression results found in Table 10[12]:

Table 10: Regression Outputs of Gallup Averages vs. BLS Unemployment Rate

	P-Value	Constant (α)	Coefficient (β)
Method 1, 1 Week	0.212	0.0131	-0.162
Method 2, 3 Week	**0.028**	**0.0178**	**-0.34**
Method 3, 5 Week	0.059	0.0095	-0.207
Method 4, EWMA	0.189	0.0152	-0.207

Although the second method has a statistically significant p-value, the coefficient is in the

wrong direction. According to this regression, when Gallup's 3 week target week average

increases by 1% the BLS's release will decrease by about .3%. Similarly, the third method (5

week average) has a p-value which is nearly statistically significant, but the coefficient is

negative. Finally, the exponentially weighted moving average has the same negative coefficient

---

[12] Regression outputs are summarized in the Appendix.

as the third method and a p-value which is substantially higher. These results are likely the result of a sample size that is simply too small.

Although these findings conclude that Gallup's data cannot forecast the BLS's unemployment rate, there are other BLS unemployment figures that might lead to a positive conclusion. Since the BLS figures are for 16 and older and Gallup's are for 18 and older, there is a clear misalignment between the two data samples that can be improved upon. The BLS also releases a 20 and older (20+) unemployment rate with no seasonal adjustment in its monthly document which can be replicated by Gallup to create a more "apples-to-apples" comparison. Although it may seem that changing Gallup's sample to remove people of age 18 and 19 is a minor adjustment, the following results prove otherwise.

Figure 11 and Table 10 show the unemployment rate forecast results of the first three methods previously explained for the standard unemployment rate when applied to the new 20+ data. These regression outputs follow the same equation relationship explained above Table 9.

Figure 11: Gallup's 20+ Unemployment Rate Averages vs. BLS 20+ Unemployment Rate
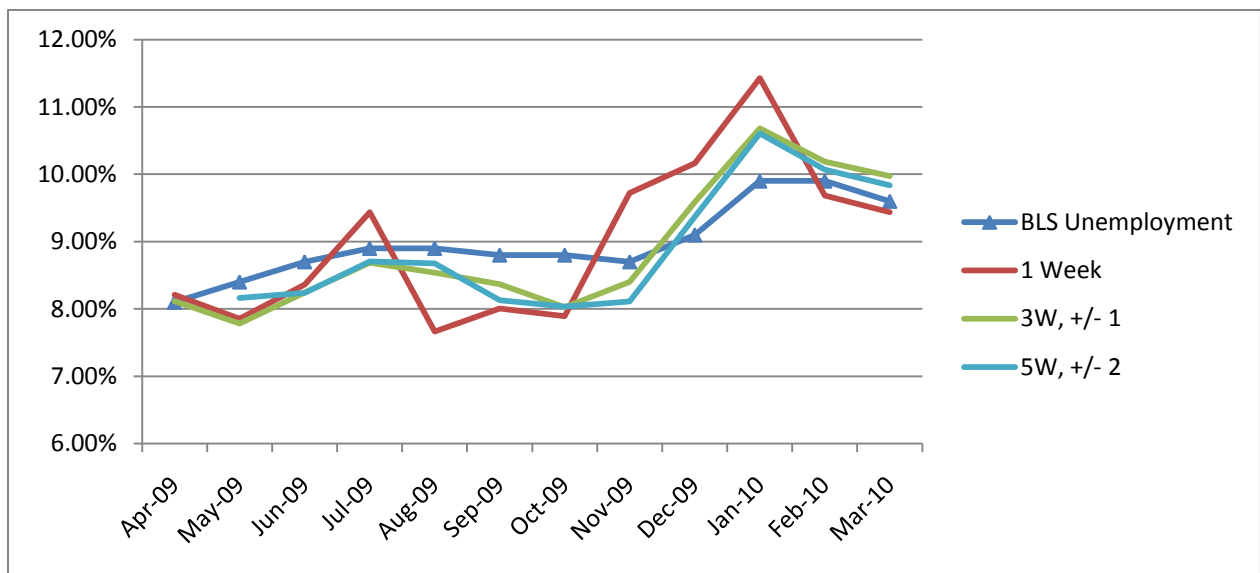
Table 11: Regression Outputs of Gallup's 20+ Unemployment Rate vs. BLS 20+

	1 Week	3W, +/- 1	**5W, +/- 2**
Constant	0.0147	0.00889	**0.005686**
Coefficient	0.0718	0.3453	**0.39272**
P-Value	0.436	0.022	**0.004**
R-Squared	6.90%	45.70%	**67.60%**

Based on Figure 11 and Table 11 one can immediately see that the results are significantly stronger for the 20+ data than for the standard unemployment rate. First of all the p-value is statistically significant for both the 3 week and 5 week numbers. Second, the coefficients are in the *right* direction. This implies that as Gallup's 20+ unemployment rate goes up, one can expect the BLS's 20+ unemployment rate to increase as well.

Despite these already strong results, by changing the averaging methodology slightly, I was able to attain a stronger relationship. If we increase the weighting of Gallup's data based on how close it is to the target week, we obtain the following two formulas for the 3 week and 5 week averages:

The λ weighting coefficient is constrained to be greater than zero and less than one and the formula operates in such a way that the closer λ is to one, the more the EWMA will resemble a simple straight moving averaging. As λ is decreased, then the weekly averages further away

from the target week are given less weight than the target week itself. This creates a two sided

exponential model which gives more weight to more relevant and appropriate weeks.

Figure 12 displays the results of this approach for both the 3 week and 5 week

exponentially weighted moving averages with λ set at .9, .8, and .7.

Figure 12: EWMA Regression Outputs for Gallup's 20+ Unemployment Rate vs. BLS's 20+
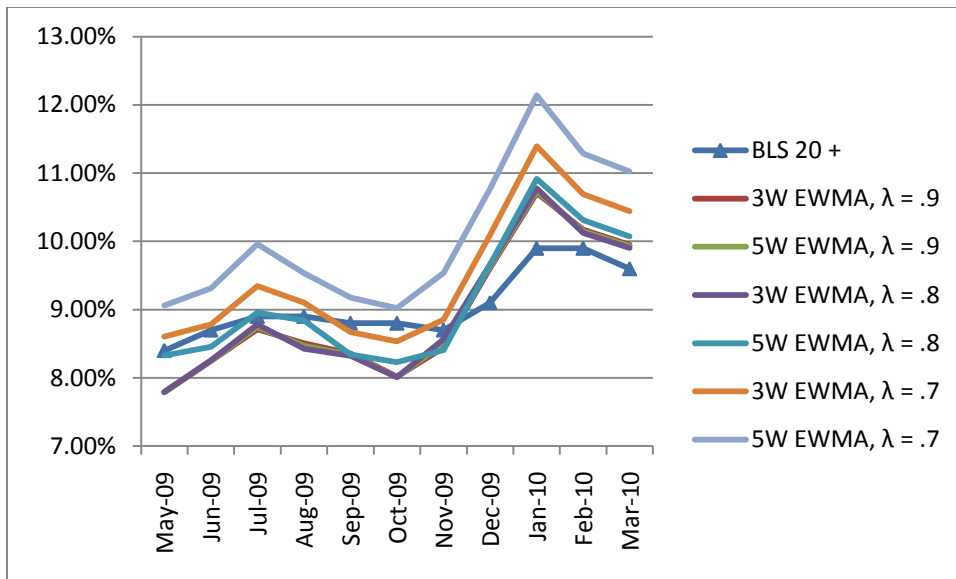
The overall trend here is that as λ is decreased, the results become more volatile and less

precise. There also doesn't appear to be a significant difference in results between the 3 week

and the 5 week average. Table 12 below summarizes the results of the regressions which follow

the same relationship as the previous set of unemployment regressions:

Table 12: EWMA Gallup's Unemployment Rate Regression Outputs

	3W EWMA, λ = .9	5W EWMA, λ = .9	3W EWMA, λ = .8	5W EWMA, λ = .8	3W EWMA, λ = .7	5W EWMA, λ = .7
Constant	0.00302	0.00545	0.003489	0.005415	0.004087	0.00578
Coefficient	0.4087	0.39876	0.3947	0.3948	0.3753	0.3718
P-Value	0.007	0.004	0.01	**0.004**	0.014	0.01
R-Squared	61.30%	62.90%	58.60%	**68.40%**	55.00%	58.50%

Once again, as opposed to the 16+ standard unemployment rate, the 20+ results consistently have positive coefficients[13]. In this particular selection of exponentially weighted moving averages and lambdas, all of the regressions turn out to be statistically significant. Furthermore, the 5 week EWMA with a λ of .8 has a lower p-value than the best straight average and also has a higher R-Squared. Since these regressions are only based on 11 periods, the results are likely to change as Gallup continues to collect its data and increase the length of its time series.

The conclusion of this second hypothesis is twofold. First, although Gallup's unemployment rate is consistently within 1.5% of the BLS's unemployment rate, this difference, coupled with a short dataset, disproves the hypothesis that Gallup's data can predict the BLS unemployment rate. In the case of the standard unemployment rate, Gallup's forecasts are quite accurate, but not accurate enough for trading purposes. Second, when the analysis is extended to study the relationship between Gallup's 20+ data and the BLS's 20+ data, the results are remarkably different. Both a straight average and a two-tailed exponentially weighted average conclude with statistically significant regressions.

It is important to note that although this analysis has proven that Gallup's data *can* forecast the BLS's 20+ unemployment rate, this rate is *not* as relevant to financial markets. The analysis conducted for the first hypothesis assesses the impact of the standard unemployment rate surprise on the S&P500. It is very unlikely that a market consensus figure exists for the 20+ unemployment rate and thus we are left with one of two assumptions. If one assumes that the surprise in the standard unemployment rate is equal to the surprise in the 20+ unemployment rate, then based on the 20+ unemployment rate surprise one can determine the standard

---

[13] I conducted this two tailed EWMA analysis for the standard unemployment rate data and the results concluded with a negative coefficient every time.

unemployment rate surprise and make an excellent trade. On the other hand, if one assumes that this is not the case, then Gallup's data cannot be used for trading purposes and a study of the relationship between the standard unemployment rate and the 20+ unemployment rate should be conducted.

7. Conclusion

We are all consumers and as such, our actions will always play an integral role in the state of our economy. Consumer data, such as the unemployment rate, is monitored by investors, analysts and executives every day in order to make educated decisions. The first hypothesis, has shown that if we know what the unemployment rate will be, there is a profitable strategy that will allow an investor to confidently profit off of large unemployment surprises. Furthermore, the first hypothesis has proven there exists a negative relationship between the unemployment rate surprise and the overnight return of the S&P500 and that this negative relationship becomes stronger as the surprise rises. The second hypothesis has shown that although Gallup's Daily Tracking unemployment rate data cannot forecast the standard 16+ unemployment rate, it can forecast the BLS's 20+ unemployment rate. Through the use of a two-sided exponentially weighted average, Gallup's data can forecast the BLS's release on a statistically significant level. It is important to note, that this data remains exploitable only as long as Gallup keeps its daily figures private. Through the fact that the BLS operates with a 3 week delay in releasing the unemployment rate after actually collecting it, Gallup's data is able to gain an informational asymmetry over the market. Gallup is able to collect crowd core data: data that is fact and not opinion.

Despite the statistically significant relationship determined in the second hypothesis, the results may prove to be quite different after a longer time series for comparison is established. Furthermore, the current unemployment situation may prove to be an outlier period relative to other periods due to the drastically changing unemployment environments.

8. FUTURE RESEARCH

There are a variety of areas of further research and this section will focus on four primary areas: emerging market, financial product specific research, real estate research and industry specific research.

An Emerging Market Approach

Based on the findings from the first two hypotheses, we can conclude that Gallup's unemployment rate can be used to forecast the BLS release (at least the 20+ figure) and that if we know the BLS's release we can implement a profitable trading strategy based on the size of the unemployment surprise. Furthermore, these two conclusions prove that Gallup's methodology determines the *true* unemployment rate given a small confidence interval. If one assumes that the BLS releases this *true* unemployment rate, then Gallup is consistently able to forecast this *true* rate with about a 1-2% confidence interval[14].

In addition to the Daily Tracking survey, Gallup also has a project titled the World Poll. The World Poll is an annual survey in every country on Earth of approximately 1000 people. Although this survey asks a variety of questions ranging from question about well-being to confidence in a variety of institutions to opinions, the World Poll collects its own estimate of what a nation's unemployment rate is. The survey uses only 1000 data points a year, which leads to wider confidence intervals than in the U.S., but I believe that valuable financial information can be collected.
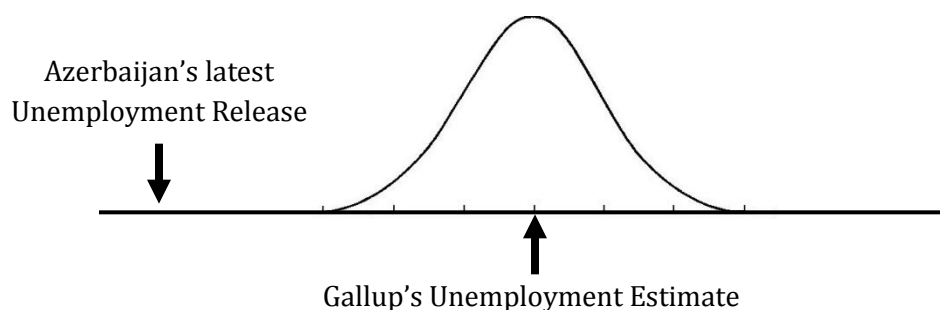
Gallup currently estimates the unemployment rate in 84 countries. Given a conservative confidence interval of approximately 5%, one can determine an estimate for what the *true*

---

[14] According to Gallup's website, its confidence interval is +/- .7% for a 30-day unemployment rate rolling average.

unemployment rate of a country should be. Therefore, if Gallup's 1000 person survey in

Azerbaijan estimates that the unemployment rate is approximately 21.1%, with 95% confidence

we can assume that the true unemployment rate is somewhere between 16.1% and 26.1%. But

the CIA World Factbook indicates that the unemployment rate in Azerbaijan is approximately

6% as of 2009[15]. Some may argue that the CIA World Factbook may often have outdated or

inaccurate unemployment data for emerging market nations, but even the International Labour

Organization (ILO) claims an unemployment rate of 6.1% (which happens to be exactly the same

as the "official estimates") as of 2008[16] . Thus the majority of economic estimates and the

financial markets will use this figure as an estimate for what *true* unemployment is in

Azerbaijan.

   But if we assume that the true unemployment rate must fall within the confidence interval

of Gallup's findings, then there is a clear contradiction. The reason for this contradiction is likely

the result of Azerbaijan purposefully underestimating its unemployment rate for a variety of

economic benefits (lower sovereign credit spreads, more foreign direct investment, etc). This is

similar to the now famous events of Argentina significantly understating its inflation rate

throughout the late 1990s. The simple visual depection below summarizes the Azerbaijan

unemployment rate scenario where the *true* unemployment rate falls somewhere below the

normal curve.

Azerbaijan's latest
Unemployment Release

Gallup's Unemployment Estimate

---

[15] https://www.cia.gov/library/publications/the-world-factbook/geos/aj.html
[16] http://laborsta.ilo.org/STP/guest

If one assumes that markets do not price in Gallup's data, which as of today is highly likely given the lack of financial clientelle that the firm has, then there is potential for arbitrage here. This arbitrage could be the result of either corrupt government data or simply the result of a lack of a more recent government or ILO estimate. If one short sells economies that understate their unempoyment rate and goes long countries who overstate their unemployment rate or countries whose unemployment rate has significantly declined, there is potentially opportunity for easy profits. This trading strategy can be implimented through a variety of financial vehicles such as currencies, sovereign debt, or even equities.

After reviewing the finding for the 84 countries that Gallup has collected its own estimates for and comparing these figures to the CIA World Factbook figures, the study shows 18 countries with an unemployment rate which are understated by the World Factbook and five countries with an unemployment rate which is overstated. Table 13 below displays the 18 countries identified sorted by which strategy to use.

Table 13: Long/Short Emerging Market Strategies

Long	Long	Long	Short
Afghanistan	Kyrgyzstan	Slovenia	Armenia
Algeria	Lithuania	Sudan	Azerbaijan
Bahrain	Malawi	Tunisia	Georgia
Djibouti	Mauritania	Turkmenistan	Palestine
Iraq	Nepal	Uganda	United Arab Emirates
Jordan	Senegal	Yemen	

Gallup has only estimated an unemployment rate on a global basis for one year and further studies should be conducted in order to investigate the implications of this strategy once more historical data is available.

Financial Product Application: Mortgage Backed Securities

In addition to emerging markets, Gallup's methodology could prove to be extremely useful for valuing mortgage backed securities. For example, let's assume that Gallup polls 1000 individuals on a daily basis and asks then about their mortgage payments. Questions could include the following:

- Are you having trouble making your mortgage payments?

- Have you missed a mortgage payment in the last month?

- If your home value drops to below your mortgage value, would you consider defaulting on your mortgage?

All three of these questions could prove to add value to a financial analyst attempting to value MBS related securities (i.e. CDO, CDO^2, CDO^3, etc). This methodology could provide analysts with a timely index of the state of the mortgage market. This index will likely be most useful for more junior tranches of CDOs. Given a large enough sample size, this information could be geographically segmented to gain an understanding of the mortgage market by state or by region. This methodology could have allowed for a more timely response to the housing crises of 2007-2008.

Real Estate Application: Migration

It is arguable that real estate prices fluctuate based on two primary reasons: capital flow and migration. Gallup's 1000 person poll could prove to be useful to forecasting real estate prices though its ability to analyze migration. If a real estate investor knows the net migration flow (inflow minus outflow) of certain cities, this data could prove to be very useful in making

investment decisions. Questions would have to gage when and where people are moving and the certainty with which they are moving.

Industry Specific Application: Crowd Core Data

Further research should assess the industry specific implications of the unemployment surprise. There are two primary approaches to industry specific applications. First, do different industries react differently to unemployment rate surprise? Does the automotive industry react differently to unemployment surprise than the career services industry? For example, Monster Worldwide, Inc.'s (MWW) share price will likely rise as a results of a positive unemployment surprise whereas the share price of Ford should decrease. Strategies such as this one may prove to have much higher Sharpe ratio due to a stock's or industry's increased sensitivity to the unemployment rate surprise relative to an entire index.

Second, industry specific research can be conducted by extracting more crowd core data through using Gallup's survey. For example, how many Americans purchased a car since the last quarterly earnings release period of major automotive companies? This approach will allow for a more accurate estimate of the revenues of an entire industry. Will more cars be sold this quarter than what market consensus estimates? These ideas can be extended to a variety of industries to gain an insider look on what is going on in the lives of consumers.
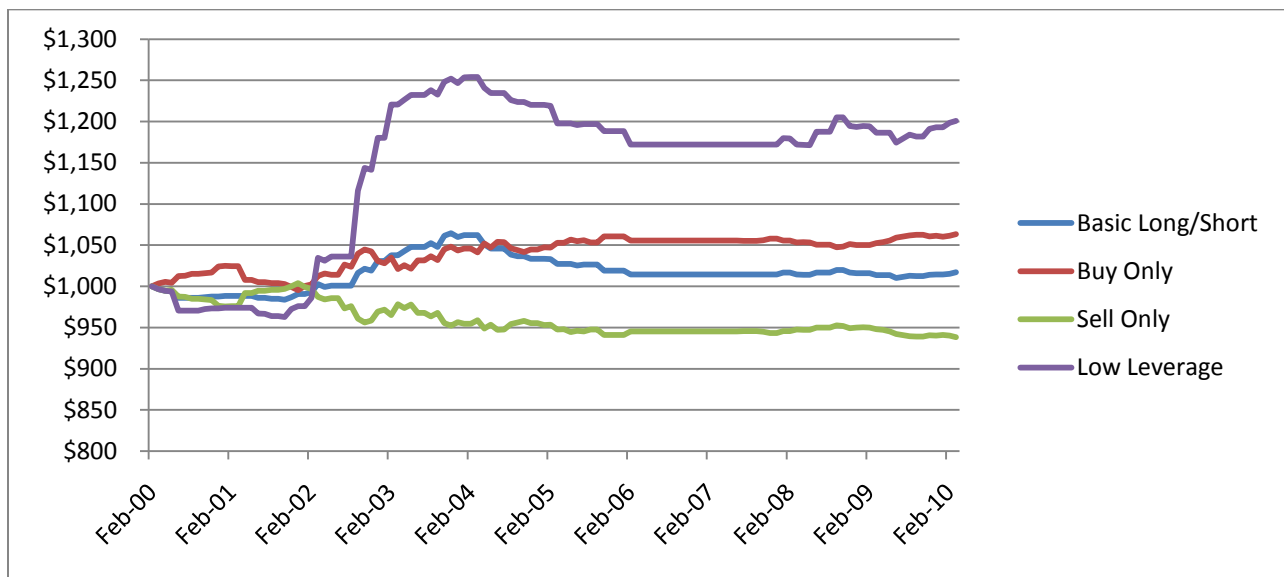
9. APPENDIX

Appendix Exhibit 1: Unemployment Rate Surprise Trading Strategies and Returns

Leverage for Each Class of Unemployment Rate Surprise

| Strategy	-0.30%	-0.20%	-0.10%	0.00%	0.10%	0.20%	0.30%	0.40%	Final Value
Basic Long/Short	1	1	1	0	-1	-1	-1	-1	1017
Buy Only	1	1	1	1	1	1	1	1	1063
Sell Only	-1	-1	-1	-1	-1	-1	-1	-1	938
Low Leverage	5	3	1	0	-1	-3	-5	-5	1201
Medium Leverage	10	3	1	0	-1	-3	-10	-10	1500
High Leverage	25	5	1	0	-1	-5	-25	-25	2728
Very High Leverage	40	10	2	0	-2	-10	-40	-60	4884
Large Surprise Only	40	0	0	0	0	0	-40	-60	5657

Exhibit 1 displays the outcomes of eight different trading strategies that apply different leverage rates to different expected surprises. For example the "Medium Leverage" strategy uses 10x leverage for +/-.3% or .4% surprises, 3x leverage for +/-.2% surprises, no leverage for +/-1% surprises and takes no position if there is no expected surprise. Appendix Exhibit 2 and Exhibit 3 display the performance of these eight strategies over the past 10 years.

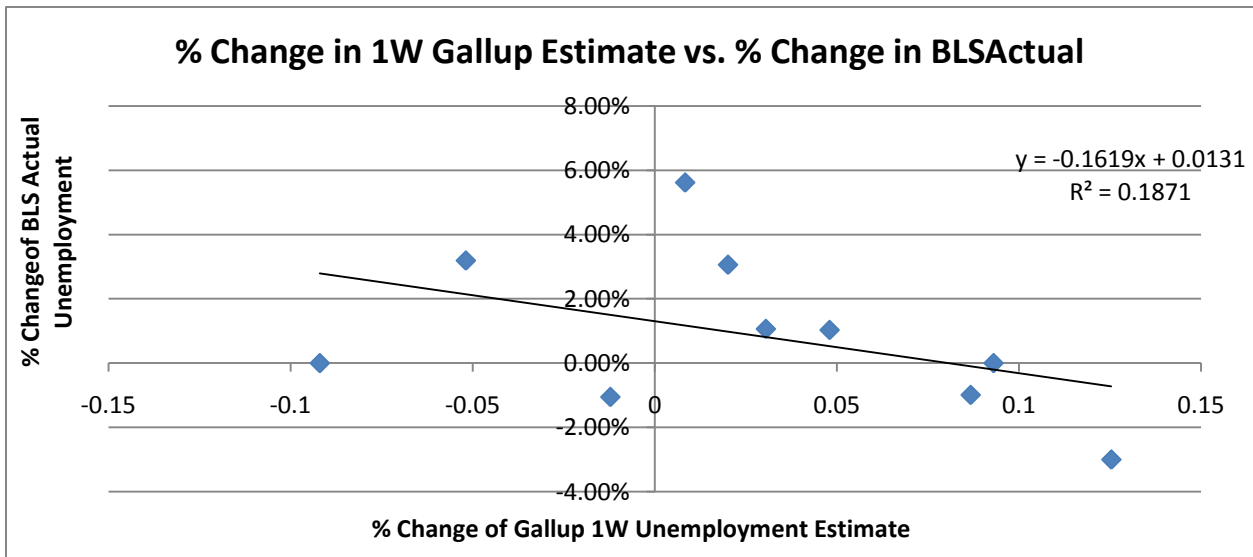Appendix Exhibit 2: Historical Performance of Low Leverage Surprise Strategies
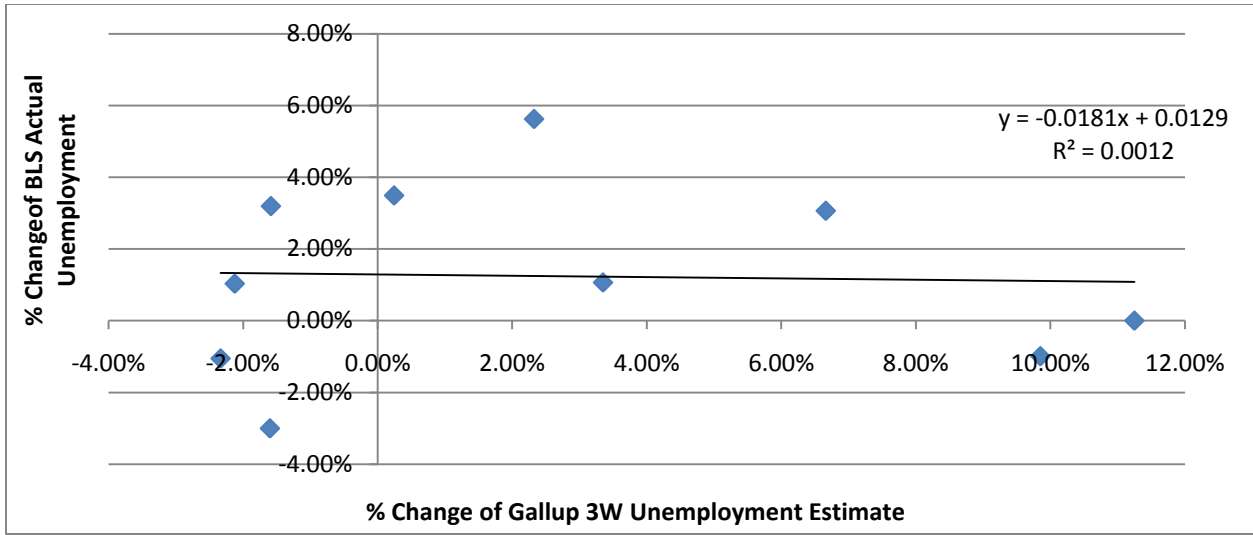
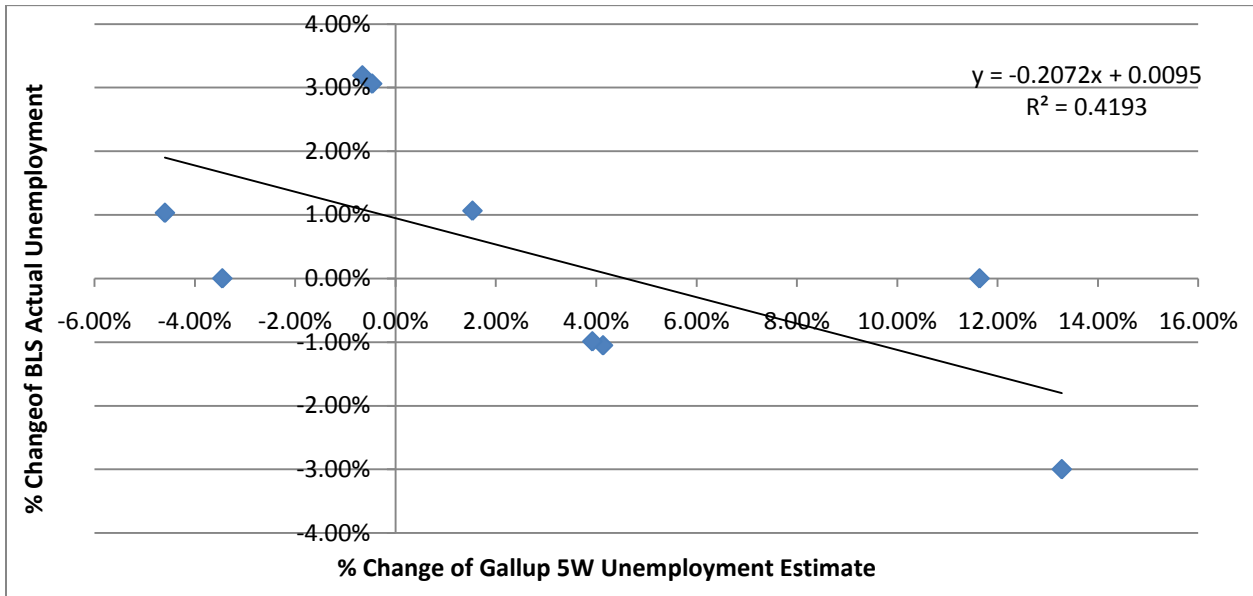Appendix Exhibit 3: Historical Performance of High Leverage Surprise Strategies
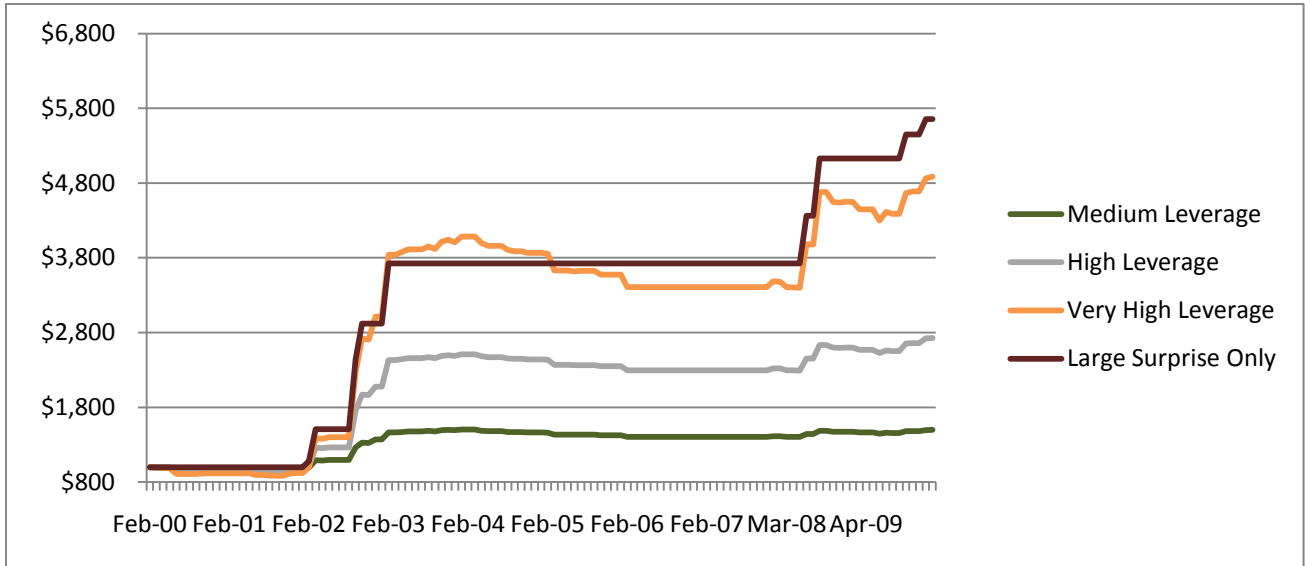
Appendix Exhibit 4: % Change Gallup 1 Week MA Estimate vs. % Change in BLS Actual

Appendix Exhibit 5: % Change Gallup 3 Week MA Estimate vs. % Change in BLS Actual

Appendix Exhibit 6: % Change Gallup 5 Week MA Estimate vs. % Change in BLS Actual

Appendix Exhibit 7: % Change Gallup EWMA Estimate vs. % Change in BLS Actual

% Change in EWMA Gallup Estimate vs. % Change in BLS Actual

$y = -0.2069x + 0.0152$
$R^2 = 0.205$
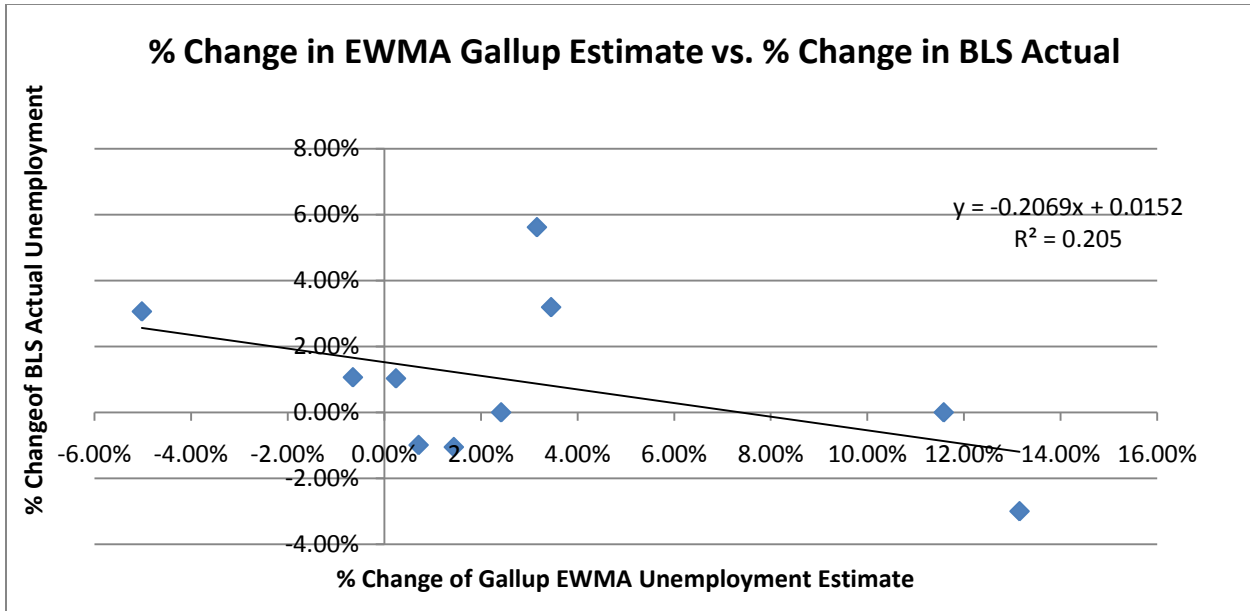
Appendix Exhibit 8: BLS 16+ Unemployment Rate and Gallup 18+ Unemployment Rate

			BLS	Gallup 18+ Unemployment Averages		
BLS Release Date	**Month**	**Representative Target Week**	**Unemployment Rate**	**Target Week 1W**	**Target Week 3W, +/- 1**	**Target Week 5W, +/- 2**
8-May-09	Apr-09	12-Apr-09	8.90%	8.26%	8.39%	
5-Jun-09	May-09	10-May-09	9.40%	8.32%	8.41%	8.44%
3-Jul-09	Jun-09	7-Jun-09	9.50%	8.58%	8.60%	8.57%
7-Aug-09	Jul-09	12-Jul-09	9.40%	8.47%	8.89%	8.93%
4-Sep-09	Aug-09	9-Aug-09	9.70%	8.03%	8.69%	8.87%
2-Oct-09	Sep-09	6-Sep-09	9.80%	8.42%	8.55%	8.46%
6-Nov-09	Oct-09	11-Oct-09	10.10%	8.59%	8.37%	8.42%
4-Dec-09	Nov-09	8-Nov-09	10.00%	9.33%	8.92%	8.75%
7-Jan-10	Dec-09	6-Dec-09	10.00%	10.20%	9.80%	9.77%
5-Feb-10	Jan-10	10-Jan-10	9.70%	11.48%	10.90%	11.07%
5-Mar-10	Feb-10	7-Feb-10	9.70%	10.42%	10.66%	10.68%

Appendix Exhibit 9: BLS 20+ Unemployment Rate and Gallup 20+ Unemployment Rate Averages

			BLS	Gallup		
Release Date	Month	Representative Target Week	Unemployment Rate	Taget Week 1W	Target Week 3W, +/- 1	Target Week 5W, +/- 2
8-May-09	Apr-09	12-Apr-09	8.10%	8.39%	8.12%	
5-Jun-09	May-09	10-May-09	8.40%	7.85%	7.78%	8.16%
3-Jul-09	Jun-09	7-Jun-09	8.70%	8.36%	8.24%	8.24%
7-Aug-09	Jul-09	12-Jul-09	8.90%	9.44%	8.69%	8.70%
4-Sep-09	Aug-09	9-Aug-09	8.90%	7.66%	8.54%	8.68%
2-Oct-09	Sep-09	6-Sep-09	8.80%	8.01%	8.37%	8.13%
6-Nov-09	Oct-09	11-Oct-09	8.80%	7.89%	8.03%	8.03%
4-Dec-09	Nov-09	8-Nov-09	8.70%	9.72%	8.40%	8.11%
7-Jan-10	Dec-09	6-Dec-09	9.10%	10.17%	9.59%	9.37%
5-Feb-10	Jan-10	10-Jan-10	9.90%	11.43%	10.68%	10.60%
5-Mar-10	Feb-10	7-Feb-10	9.90%	9.68%	10.19%	10.07%
2-Apr-10	Mar-10	7-Mar-10	9.60%	9.43%	9.97%	9.84%

Appendix Exhibit 10: BLS 20+ Unemployment Rate and Gallup 20+ Unemployment Rate EWMA

Date	BLS 20 +	3W EWMA, $\lambda = .9$	5W EWMA, $\lambda = .9$	3W EWMA, $\lambda = .8$	5W EWMA, $\lambda = .8$	3W EWMA, $\lambda = .7$	5W EWMA, $\lambda = .7$
10-May-09	8.40%	7.79%	7.79%	7.79%	8.33%	8.61%	9.06%
7-Jun-09	8.70%	8.25%	8.25%	8.26%	8.45%	8.78%	9.31%
12-Jul-09	8.90%	8.71%	8.75%	8.78%	8.96%	9.35%	9.96%
9-Aug-09	8.90%	8.51%	8.47%	8.43%	8.84%	9.10%	9.53%
6-Sep-09	8.80%	8.35%	8.34%	8.32%	8.34%	8.67%	9.18%
11-Oct-09	8.80%	8.02%	8.01%	8.01%	8.23%	8.53%	9.02%
8-Nov-09	8.70%	8.45%	8.50%	8.56%	8.41%	8.86%	9.54%
6-Dec-09	9.10%	9.61%	9.63%	9.66%	9.66%	10.09%	10.77%
10-Jan-10	9.90%	10.71%	10.74%	10.77%	10.91%	11.39%	12.14%
7-Feb-10	9.90%	10.17%	10.15%	10.12%	10.31%	10.69%	11.29%
7-Mar-10	9.60%	9.95%	9.93%	9.90%	10.07%	10.44%	11.02%

Appendix Exhibit 11: Table 8 and 9 Regression Output Averages by Lag
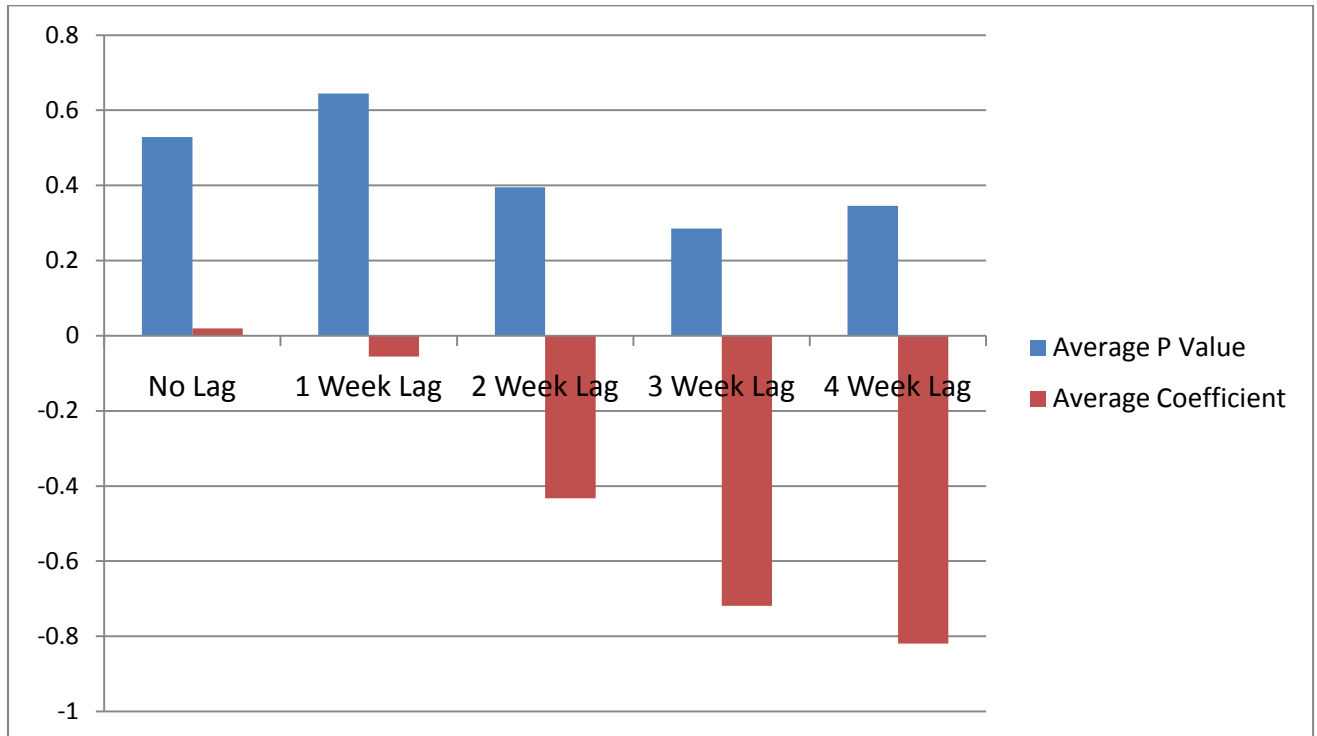
10.   END NOTES AND REFERENCES

"How the Government Measures Unemployment." *U.S. Bureau of Labor Statistics*. Bureau of Labor Statistics, 16 Oct. 2009. Web. 1 Apr. 2010. <http://www.bls.gov/cps/cps_htgm.htm#why>.

"Qualitative Call with Gallup." Telephone interview. 20 Feb. 2010.

Boyd, John H., Jian Hu, and Ravi Jagannathan. "• The Stock Market's Reaction to Unemployment News: Why Bad News Is Usually Good for Stocks." *The Journal of Finance* 60.2 (2005): 649-72. *Blackwell Publishing for the American Finance Association*. Web. 10 Apr. 2010. <http://www.jstor.org/stable/3694763>.

Surowiecki, James. *The Wisdom of Crowds*. New York: Anchor, 2005. Print.