Journal of Agribusiness 21,1 (Spring 2003):39–64

The Theory of Contrary Opinion: A Test Using Sentiment Indices in Futures Markets

Dwight R. Sanders, Scott H. Irwin, and Raymond M. Leuthold

The theory of contrary opinion predicts price reversals following extremes in market sentiment. This research tests a survey-based sentiment index's usefulness as a contrary indicator across 28 U.S. futures markets. Using rigorous time-series tests, the sentiment index displays only a sporadic and marginal ability to predict returns, and in those instances the pattern is one of return continuation—not reversals. Therefore, futures traders who rely solely upon sentiment indices as contrary indicators may be misguided.

Key Words: bullish consensus, contrary opinion, market sentiment

> *Market sentiment can be an invaluable tool when it comes to picking market turning points. When sentiment readings reach an extreme it gives you an alert to a possible turn in the market. It signals an imbalance in the market; if 90% of traders are bullish at the end of the day, who is left to buy? The futures market lends itself to be an ideal market for this type of analysis.*
>
> — *Daily % Bullish*, October 2000

For many years, sentiment has been widely used as a contrary opinion indicator in futures markets (Neill, 1960). Trading based on the "theory of contrary opinion" generally is defined as taking a market position that is opposite of the prevailing market opinion or psychology, and it is considered to be a "solidly logical" technical approach to trading futures (Teweles and Jones, 1999, p. 179). Ironically, the popularity of the approach has led to the development of survey-based sentiment indices, which directly measure the level of agreement among a segment of market participants. When these sentiment indices reveal a "predominant number of market analysts are bullish [bearish], it is quite likely that the market is approaching an overbought [oversold] condition, and that a reversal in trend may be imminent" (Consensus, Inc., 2001). As highlighted in the opening quote, futures traders frequently rely on

Dwight R. Sanders is assistant professor, Department of Agribusiness Economics, Southern Illinois University, Carbondale; Scott H. Irwin and Raymond M. Leuthold are professors, both in the Department of Agricultural and Consumer Economics, University of Illinois, Urbana.

sentiment indices as a measure of market opinion and use them to make trading decisions under the theory of contrary opinion.

Market sentiment can be measured either indirectly through market-based sentiment indicators, or directly through surveys of market participants. An example of an indirect, or market-based, measure of market sentiment is the put-call ratio: the total trading volume in puts divided by the total trading volume in calls. Simon and Wiggins (2001) refer to the put-call ratio as a "fear indicator." When the ratio is at an extremely high level, then market participants are more active in puts; thus, they are bearish or fearful of a market decline. Under the theory of contrary opinion, this would portend a market rally. An example of a direct, or survey-based, measure of market sentiment is the Bullish Sentiment Index published by *Investors Intelligence*. This sentiment index is based on a survey of stock market newsletter writers, and reflects the level of agreement of the writers about the market outlook (Clarke and Statman, 1998). Under the theory of contrary opinion, if a majority of newsletter writers are bullish, then the market is overbought and it is expected to decline, and vice versa.

There is clearly a group of traders who think market sentiment is an important indicator for predicting futures prices. Yet, despite its widespread use in futures trading, few studies have examined the ability of sentiment measures to predict futures market returns. Prior research generally has focused on stock markets (Solt and Statman, 1988; Neal and Wheatley, 1998; Brown and Cliff, 2001). The limited research on futures markets has relied on indirect measures of market sentiment (Wang, 2001; Simon and Wiggins, 2001). There is an apparent gap in the literature with regard to the usefulness of direct, or survey-based, sentiment indices as contrarian indicators in futures markets. The purpose of this study is to test the predictability of futures returns using a direct measure of market sentiment—the Consensus Index of Bullish Market Opinion published by Consensus, Inc.

In addition to the use of a direct measure of sentiment, this research expands the existing literature in three other ways. First, the research is comprehensive in that it examines a total of 28 futures markets. Second, care is taken to fully investigate and present the behavior of the sentiment indicators, the opinions they capture, and how they are compiled. Third, a rigorous time-series methodology is employed to test the relationship between the level of sentiment and the movement of subsequent futures returns.

The following section reviews previous studies on the usefulness of sentiment in predicting market returns. The next section introduces the Consensus Index of Bullish Market Opinion and presents a thorough description of the data set. This is followed by an explanation of the methodology and the results. The paper concludes with a summary and discussion of the results and possible ramifications for academics and practitioners.

Previous Studies

Researchers who have examined the predictability of stock returns using market sentiment have reported mixed results. Neal and Wheatley (1998) found that the odd-lot ratio—an indirect measure of public participation in the market—does not predict market returns. Similarly, Elton, Gruber, and Busse (1998) found no evidence that small investor sentiment, as measured by the discounts on closed-end funds, is an important factor in the return generating process.

Solt and Statman (1988) examined the sentiment of retail stock investors as captured in the Bearish Sentiment Index compiled by *Investors Intelligence*. This is a direct gauge of market sentiment which is constructed by surveying market newsletters as to their outlook. Solt and Statman concluded that this market sentiment index contains no useful information for forecasting market returns. Using the same data set, Clarke and Statman (1998) confirmed the sentiment of newsletter writers does not forecast equity returns, but past returns and market volatility do affect sentiment.

Using measures of sentiment among Wall Street strategists (stock allocation recommendations), newsletter writers (Bullish Sentiment Index), and individual investors (AAII survey), Fisher and Statman (2000) found investor sentiment can differ across groups. They show that the levels of sentiment among Wall Street strategists and individual investors are reliable contrary indicators of market direction, but there is not a statistically significant forecasting relationship between the Bullish Sentiment Index and the stock market. Fisher and Statman specifically encourage additional research on market sentiment in other markets with other sentiment measures.

Simon and Wiggins (2001) investigated the usefulness of market-based sentiment indicators in the S&P 500 futures market. In their study, market sentiment is measured with the volatility index (implied volatility from the S&P 100 index options), the put-call ratio (total volume of puts divided by total volume of calls traded on the S&P 100 options), and the trading index or TRIN (a scaled measure of number of advancing stocks divided by the number of declining stocks). The authors found that the sentiment indicators are statistically and economically useful contrarian indicators in the S&P 500 futures market—i.e., a high level of bearishness or fear in the stock market leads to subsequent positive returns in the S&P 500 futures.

Wang (2001) examined the impact of market participant sentiment in agricultural futures markets using Commodity Futures Trading Commission (CFTC) *Commitment of Traders* reports to gauge the sentiment of reporting noncommercials (large speculators), reporting commercials (large hedgers), and nonreporting traders (small traders).[1] Based on Wang's findings, large speculator sentiment predicts price continuation, large hedger sentiment predicts price reversals, and small trader sentiment is not useful in predicting prices. However, the returns to large speculators appear to be a premium for absorbing hedging pressure, and are not due to superior forecasting skills.

[1] It is not clear if the CFTC *Commitment of Traders* data represent a direct or indirect measure of market sentiment. While these data are clearly not market-based, neither are they truly survey-based. Rather, the data are based on a predefined classification system. Although actual positions are represented, the motivations for these positions, especially among "hedgers," is not known.

The research presented in the current study differs from and expands that of Simon and Wiggins (2001) and Wang (2001) by utilizing a direct, or survey-based, measure of sentiment. In a related line of research, Sanders, Irwin, and Leuthold (2000) used a survey-based measure of market sentiment (Market Vane's Bullish Sentiment Index) as a proxy for noise trader sentiment to directly test the predictions of a theoretical noise trader model. Here, a different sentiment index (the Consensus Index of Bullish Market Opinion) is employed, and the focus is on testing the theory of contrary opinion. Specifically, we address the question: Are survey-based sentiment indices useful in predicting returns in futures markets?

Measuring Market Sentiment

The methodology used by Consensus, Inc. to compile its bullish sentiment index is quite simple. Consensus, Inc. publishes a weekly market paper, *CONSENSUS: National Futures and Financial Weekly*, containing a sampling of investment newsletters. From the sample of letters Consensus, Inc. receives, it compiles a sentiment index with a simple count of the number of bullish newsletters as a proportion of all newsletters expressing an opinion. Consensus, Inc. considers only those opinions committed to publication. The Consensus Bullish Sentiment Index at time t ($CBSI_t$) is expressed as:

$$CBSI_t = \frac{\textit{Number of Bullish Newsletters}}{\textit{Number of Newsletters Expressing an Opinion}}.$$

For instance, if Consensus, Inc. receives 100 newsletters commenting on the frozen pork bellies market, and 25 of those think that pork belly prices are going to increase, then the CBSI is 0.25, or 25%.[2] The CBSI is compiled each Friday, reflecting the opinions expressed in newsletters which are published during the week. It is released early the following week by recorded telephone message and published in the following Friday's edition of *CONSENSUS*.

Fisher and Statman (2000) note that the sentiment of different trading groups (e.g., Wall Street strategists versus individual investors) can provide distinctly different market signals. Therefore, it is important to understand how indices are compiled, the types of information used by survey participants, and the group of traders who may be acting upon their advice. Here, we carefully examine these issues to aid in our understanding of the data, to facilitate comparisons with other research, and to assist in the interpretation of the results. Consensus, Inc. surveys newsletter writers in futures markets. But, what information sources do the newsletter writers utilize in forming their market opinions, and what group of traders is acting upon that advice? A brief review of the decision-making rules of small traders and a sampling of their information sources help address these questions.

---

[2] Consensus, Inc. acknowledges some interpretation is required for newsletters which do not explicitly make buy or sell recommendations.

In an early study, Smidt (1965) documented that most amateur speculators surveyed preferred to trade commodities about which they had personal knowledge or advice. Surveys by the Chicago Board of Trade and *Barron's* have reported similar findings (see also Nagy and Obenberger, 1994; Brennan, 1995). As summarized by Draper (1985), the surveys suggest the average futures trader's sources of information include: articles/publications, broker and newsletter recommendations, advisory services, and self-analysis. Consistent with these findings, Canoles et al.'s (1998) survey of retail futures traders reveals their favorite sources of information are professional trading advisory services and general financial publications. Collectively, these results indicate retail speculators collect much of their information from focused media sources, such as those surveyed by Consensus, Inc.

Based on the literature reviewed above, market advisors, brokers, and newsletters provide decision-making information for retail futures speculators. But, how do newsletter writers form their market opinion? Two excerpts from the February 17, 1995, issue of *CONSENSUS: National Futures and Financial Weekly* (Consensus, Inc.) provide insight as to the information contained within advisors' recommendations and market newsletters:

> The major uptrending channel line is at 102-00 today. The strong close puts the market in a strong position once again. The old main top at 102-29 was taken out. This means that 101-08 is the new main bottom. Now that the (T-Bond) market has closed inside of the uptrending channel the upside potential is 103-17. Long-term swing chart is still projecting a rally to 103-26 by February 24th [contributed by James A. Hyerczyk, Hyerczyk Technical Comments].

Each issue of *CONSENSUS* is filled with this type of technical commentary for nearly every futures market. Many market advisors rely on technical indicators and simply pass along this information to their retail subscribers. Although less common than technical analysis, some newsletters are fundamental in nature, relaying government reports, seasonal tendencies, and pertinent cash market conditions:

> The USDA left the 1994–95 ending stocks of soybeans unchanged at 510 M.B. which suggests that the market will not be as sensitive to weather as corn or possibly wheat.... Seasonally, the market tends to bottom in late February and work higher into March and May [contributed by Strickler, Bradford & Co., Inc.].

While the newsletters often contain detailed interpretations of relevant supply and demand factors, the fundamental analysis tends to reiterate public information. In aggregate, the surveyed newsletters seem to rely heavily on technical analysis, and to a lesser degree on fundamental analysis, in forming their market opinions.

In summary, Consensus, Inc. surveys market newsletter writers, and retail speculators appear to be the typical audience for this printed material. So, while the CBSI reflects sentiment among newsletter writers, the information is likely acted upon by retail speculators. This connection is consistent with the relatively high level of correlation between newsletter writer sentiment and individual investor sentiment reported by Fisher and Statman (2000) in stock markets.

The previous discussion indicates the CBSI is a valid direct, or survey-based, measure of the market opinion of newsletter writers in futures markets. However, there is some disagreement in the literature as to the applicability of survey-based versus market-based measures of sentiment. Brown (1999) supports the use of direct measures of market sentiment based on Occam's Razor—the simpler the explanation, the better. In contrast, Simon and Wiggins (2001) are critical of survey-based measures because they may be "stale" by the time they reach publication, they tend to equally weight respondents, and there is no accounting for the degree of bullishness among respondents.

In the following analysis, the data sets are carefully aligned to avoid staleness problems, yet at the same time not confounding correlation and causality among sentiment and returns. Also, the CBSI is highly correlated with Market Vane's survey-based measure of market sentiment, which does use a weighting scheme based on the degree of bullishness and the newsletter's perceived influence.[3] Thus, the CBSI appears to be an accurate and simple survey-based measure of market sentiment. In the following sections, we present the time-series data and methodology to test if the CBSI is a useful contrary indicator in futures markets.

Data

Futures Data and Markets

Weekly futures returns are calculated for the closest to expiration contract where the maturity month has not been entered. The time series of futures returns are created to match up with the sentiment data. Specifically, nearby contract returns are calculated on a Friday-to-Friday basis using closing prices. The return series corresponds to the Friday compilation of the Consensus, Inc. sentiment data.[4] Returns (R_t) are calculated as the continuously compounded change in closing prices, $\ln(p_t/p_{t-1})$. Weekly data from May 1983 through September 1994 are available, but 54 weeks are withheld for potential out-of-sample testing, which results in 536 observations.

A cross-section of 28 U.S. futures markets is examined to avoid erroneous implications based on the nuances of a particular market. Markets are chosen based on the availability of the futures and sentiment data. To facilitate the presentation of results and for relevant comparisons, related markets are designated into groups. Group classification is based on common production/consumption patterns and expectations concerning the correlation of returns and sentiment among the markets. The five

[3] Sanders, Irwin, and Leuthold (2000) present summary statistics for Market Vane's sentiment indices. Correlation results for the Market Vane and CBSI indices are available from the authors upon request.

[4] More specifically, futures returns are matched to the date the CBSI is compiled, rather than the publication date. There are two reasons why returns are matched to the compilation date. First, this minimizes "staleness" problems with Consensus information. Second, the CBSI is published in two forms during the "release week" (the week that follows the Friday date of compilation). In the early part of the "release week," the CBSI is made available to subscribers by recorded telephone message. On Friday of the "release week," it is published in the weekly edition of *CONSENSUS*. Hence, it is not possible to pinpoint a specific date that the CBSI is available to market participants.

Table 1. Summary Statistics, Consensus Bullish Sentiment Index (May 1983–September 1994)

Market [a]	Mean	Std. Dev.	Min.	Max.	Correlation [b]
Grain:					
Corn	45.701	19.916	5	92	0.289
Wheat	46.413	20.193	3	91	0.295
Soybeans	46.783	17.882	12	90	0.278
Soybean Meal	42.501	20.012	5	95	0.347
Soybean Oil	43.992	21.861	5	96	0.323
Livestock:					
Live Cattle	51.584	15.547	15	87	0.319
Feeder Cattle	46.998	19.617	6	95	0.374
Live Hogs	44.332	15.696	13	88	0.289
Pork Bellies	39.716	17.913	4	88	0.335
Food/Fiber:					
Coffee	43.992	20.906	5	96	0.286
Sugar	51.279	22.112	5	94	0.279
Cocoa	41.755	20.455	4	94	0.223
Orange Juice	40.294	22.731	6	94	0.363
Cotton	45.981	21.331	7	96	0.321
Lumber	42.181	21.033	5	94	0.328
Financial:					
Deutsche Mark	46.876	21.822	4	89	0.316
Swiss Franc	45.205	21.739	3	94	0.301
Japanese Yen	42.701	20.821	3	91	0.312
British Pound	42.870	22.017	0	96	0.273
Canadian Dollar	41.591	19.899	0	92	0.326
Treasury Bills	46.619	20.917	5	93	0.233
Treasury Bonds	44.406	17.525	9	86	0.274
Metal/Energy:					
Gold	43.570	20.630	3	96	0.233
Silver	43.531	19.254	4	95	0.203
Platinum	44.450	21.641	6	95	0.264
Heating Oil	39.679	20.469	4	87	0.270
Crude Oil	40.401	18.471	3	86	0.300
Gasoline	38.551	20.674	5	93	0.313

[a] All of the markets have 591 weekly observations, except crude oil and gasoline, which begin in April 1985 and have 494 observations.

[b] The contemporaneous correlation coefficient between market returns and sentiment. The standard error of the estimated correlations is $(1/(n-3))^{1/2}$, so with $n = 591$, the standard error is 0.04123, and any correlation coefficient greater than 0.0809 (0.106) is statistically different from zero at the 5% (1%) level using a two-tailed t-test.

groups include: grain (corn, wheat, soybeans, soybean meal, and soybean oil); livestock (live cattle, feeder cattle, live hogs, and frozen pork bellies); food/fiber (coffee, sugar, cocoa, orange juice, cotton, and lumber); financial (Deutsche mark, Swiss franc, Japanese yen, British pound, Canadian dollar, Treasury bills, and Treasury bonds); and metal/energy (gold, silver, platinum, heating oil, crude oil, and gasoline). A complete listing of markets and their summary statistics are reported in table 1.

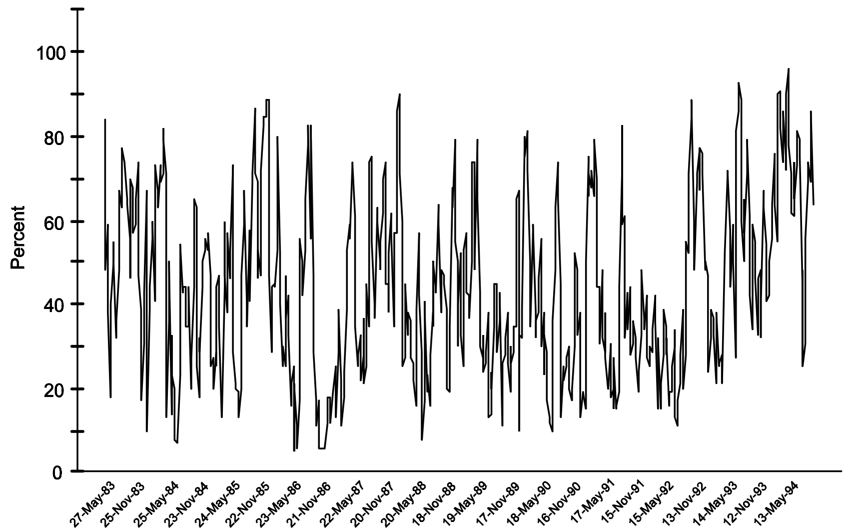

Figure 1. Consensus Bullish Sentiment Index, coffee (May 1983–September 1994)

Summary Statistics

It is necessary to examine the simple summary statistics to fully understand the data and to motivate the time-series approach used in the analysis. The general characteristics of the sentiment indices are explored with simple summary statistics presented in table 1. The mean sentiment level (% bullish) is notably less than a neutral 50 for the CBSI. In fact, the mean CBSI is statistically less than 50 at the 1% level (two-tailed *t*-test) for all the markets except live cattle and sugar. The range of the mean CBSI is from a low of 38.6 for gasoline to a high of 51.6 for live cattle.

For all markets, sentiment is quite volatile, with large standard deviations and extremes of above 90 and below 10. The extreme values of sentiment along with its volatility may suggest the newsletter writers who make up the indices are reacting to correlated market signals. As an illustration of the sentiment behavior over time, the CBSI for coffee is plotted in figure 1.

The last column of table 1 shows the contemporaneous correlation coefficient between returns and sentiment. The largest correlation is 0.374 for feeder cattle and the lowest is 0.203 for silver. It is noteworthy that the correlations are all significantly positive at the 1% level (two-tailed *t*-test). Newsletter opinions collected during the week are positively correlated with market returns during the same week. Therefore, using contemporaneous sentiment and returns in the time-series analysis could result in the erroneous conclusion that sentiment "causes" returns, when in fact just the opposite may be true. This would be a classic example of confusing correlation with causality.

Table 2 presents the simple contemporaneous correlation coefficients for sentiment and futures returns across related markets. Not surprisingly, sentiment among related commodities is highly correlated. Looking at panel A (grain), a correlation of 0.631 indicates that when newsletter writers are bullish about corn, they are also bullish about the price of soybeans. Likewise, in panel D (financial), newsletter writers tend to have similar sentiment about the price of the various currencies versus the U.S. dollar. These correlations are consistent with common decision-making factors among newsletter writers. Note, however, the correlation among sentiment indices is much lower in relatively unrelated markets, such as those in the food/fiber group in panel C.

Generally speaking, within closely related market groups, the level of correlation among sentiment indices is comparable to the correlation of futures returns across the same markets (lower diagonal entries in table 2). The relatively strong levels of correlation among both the returns and sentiment within designated market groups help to motivate and justify the pooling procedures implemented in the following sections.

Methodology and Results

Market Sentiment and Returns

Understanding the behavior of sentiment is important in examining its usefulness as a market indicator. A general method of exploring the linear linkages between sentiment and price is the "Granger causality" framework. To assure the time-series tests for return predictability are properly specified, it is important to test for causal linkages in both directions. Hamilton (1994, p. 302) suggests the following direct, or bivariate, Granger test:

$$(1) \qquad \rho_t = c_0 + \sum_{i=1}^{p} a_i \rho_{t-i} + \sum_{j=1}^{q} b_j R_{t-j} + e_t,$$

where ρ_t and R_t represent noise trader sentiment and futures returns, respectively, and e_t is a white noise error term.

Causality from returns to sentiment in equation (1) is tested under the null of $b_j = 0, \forall j$. Specifically, equation (1) is estimated with ordinary least squares (OLS), and the null hypothesis that R_t does not lead ρ_t is tested with a Chi-squared test (Hamilton, 1994, p. 305).[5,6] The aggregate sign of causality is addressed by summing

[5] Note, misspecification of equation (1) due to cointegration and an omitted error-correction term is not a problem with these data as sentiment is clearly stationary I(0) in levels.

[6] The causality test assumes that the two series, ρ_t and R_t, are covariance stationary, and e_t is an i.i.d. white noise error. This assumption is tested using White's general test for heteroskedasticity in the error term. If e_t is heteroskedastic, then the model is reestimated using White's heteroskedastic consistent covariance estimator, and the appropriate test for the parameter restrictions is a Wald Chi-squared test (Greene, 1993, p. 392). A Lagrange multiplier test is used to verify that the residuals are serially uncorrelated. If, after choosing the optimal lag length, the residuals demonstrate autocorrelation, then additional lags of the dependent variable are added as explanatory variables [i.e., p is increased in equation (1)] until the autocorrelation is eliminated.

Table 2. Correlation Matrices, Sentiment and Futures Returns Across Markets (May 1983–September 1994)

[The upper off-diagonal entries are correlations for Consensus sentiment; the lower off-diagonal entries are correlations among futures returns.]

PANEL A. GRAIN

	Simple Correlation Coefficients				
Market	Corn	Wheat	Soybeans	Soybean Meal	Soybean Oil
Corn		0.472	0.631	0.481	0.549
Wheat	0.467		0.387	0.335	0.352
Soybeans	0.691	0.362		0.692	0.693
Soybean Meal	0.594	0.314	0.868		0.332
Soybean Oil	0.575	0.319	0.776	0.461	

PANEL B. LIVESTOCK

	Simple Correlation Coefficients			
Market	Live Cattle	Feeder Cattle	Live Hogs	Pork Bellies
Live Cattle		0.673	0.470	0.268
Feeder Cattle	0.812		0.315	0.180
Live Hogs	0.440	0.369		0.654
Pork Bellies	0.245	0.231	0.597	

PANEL C. FOOD/FIBER

	Simple Correlation Coefficients					
Market	Coffee	Sugar	Cocoa	Orange Juice	Cotton	Lumber
Coffee		0.005	0.249	0.023	0.102	0.049
Sugar	0.013		0.062	0.037	0.073	0.069
Cocoa	0.159	0.060		0.006	0.046	!0.017
Orange Juice	!0.055	0.008	0.039		!0.072	!0.021
Cotton	!0.012	0.041	0.094	!0.069		0.217
Lumber	0.079	0.065	!0.035	!0.063	0.066	

PANEL D. FINANCIAL

	Simple Correlation Coefficients						
Market	Deutsche Mark	Swiss Franc	Japanese Yen	British Pound	Canadian Dollar	Treasury Bills	Treasury Bonds
Deutsche Mark		0.916	0.613	0.774	0.299	0.168	0.259
Swiss Franc	0.947		0.605	0.789	0.288	0.135	0.186
Japanese Yen	0.646	0.650		0.591	0.286	0.181	0.126
British Pound	0.769	0.775	0.488		0.331	0.134	0.152
Canadian Dollar	0.079	0.084	0.047	0.159		0.046	0.191
Treasury Bills	0.136	0.125	0.063	0.098	0.054		0.627
Treasury Bonds	0.101	0.081	0.031	0.047	0.048	0.682	

(continued . . .)

Table 2. Continued

[The upper off-diagonal entries are correlations for Consensus sentiment; the lower off-diagonal entries are correlations among futures returns.]

PANEL E. METAL/ENERGY

	Simple Correlation Coefficients					
Market	Gold	Silver	Platinum	Heating Oil	Crude Oil	Gasoline
Gold		0.700	0.611	0.101	0.087	!0.081
Silver	0.734		0.653	0.059	0.032	0.024
Platinum	0.692	0.625		0.086	0.122	0.068
Heating Oil	0.284	0.156	0.187		0.762	0.634
Crude Oil	0.287	0.164	0.185	0.887		0.751
Gasoline	0.267	0.134	0.187	0.839	0.872	

Notes: The correlations are calculated over 591 observations, except for those using crude oil and gasoline data, which begin April 5, 1985, and have 494 observations. The standard error of the estimated correlations is $(1/(n-3))^{1/2}$, so with $n = 591$, the standard error is 0.04123, and any correlation coefficient greater than 0.0809 (0.106) is statistically different from zero at the 5% (1%) level using a two-tailed t-test.

the impact of lagged returns, Σb_j, and testing if it equals zero using a two-tailed t-test. If $\Sigma b_j > 0$, then market sentiment is an increasing function of past prices.

Choosing the appropriate lag lengths (p, q) is of practical significance in performing the causality test (see Thorton and Batten, 1985; Jones, 1989). As suggested by Beveridge and Oickle (1994), the order of an autoregressive system may be best determined by searching all possible lags for the combination that minimizes a model selection criterion. For example, in (1) the model is estimated by varying the own-lag length of ρ_t from $p = 1, 2, ..., p^{max}$, and the lag length of R_t from $q = 1, 2, ..., q^{max}$, such that a total of ($p^{max} \times q^{max}$) regressions are estimated. The p, q lag length combination that minimizes Akaike's information criterion (AIC) is chosen as the final model specification. This purely objective procedure has the advantage of not placing the artificial restriction that $p = q$. Additionally, it eliminates the uncertainty in multivariate cases of deciding the order in which to enter additional variables into a model. For equation (1), all possible lag-length combinations are estimated with $p^{max} = q^{max} = 8$, and p, q is chosen to minimize AIC.

The estimation results for each market are presented in table 3. The results indicate market returns lead sentiment and the cumulative impact is positive. In each market examined, the null hypothesis that returns do not lead sentiment is rejected at the 0.01 level. The additive effect of lagged returns is statistically positive (1% level) for every market in the data set. Past returns and sentiment levels explain a fairly large portion of the variation in sentiment, with the adjusted R^2 ranging from 0.531 (feeder cattle) to 0.795 (gold) in the CBSI models. These results are consistent with prior work on sentiment (Solt and Statman, 1988; De Bondt, 1993) and conjectures that newsletter writers are often trend-followers (Clarke and Statman, 1998).

For a more general characterization of market opinion, the causality test in (1) is estimated by pooling the time-series data across the designated futures groups. The

Table 3. Granger Causality Test for Individual Futures Markets, Returns Lead Sentiment (May 1983–September 1994)

$$\left[\text{Equation (1) causality test: } \rho_t = c_0 + \sum_{i=1}^{p} a_i \rho_{t-i} + \sum_{j=1}^{q} b_j R_{t-j} + e_t\right]$$

Market	p, q	$\chi^2_{[q]}$	p-Value	$\Sigma\, b_j$	t-Statistic	p-Value	Adjust. R^2
Grain:							
Corn	1, 2	39.56	0.000	152.6	4.94	0.000	0.761
Wheat	1, 1	63.83	0.000	140.7	7.98	0.000	0.741
Soybeans	2, 2	23.70	0.000	135.3	4.17	0.000	0.701
Soybean Meal	1, 2	42.64	0.000	172.9	5.45	0.000	0.658
Soybean Oil	2, 2	45.70	0.000	178.5	6.29	0.000	0.653
Livestock:							
Live Cattle	1, 6	73.92	0.000	424.3	5.67	0.000	0.608
Feeder Cattle	4, 1	43.17	0.000	266.1	6.57	0.000	0.531
Live Hogs	2, 2	89.65	0.000	183.8	3.96	0.000	0.675
Pork Bellies	2, 3	54.17	0.000	79.3	3.96	0.000	0.630
Food/Fiber:							
Coffee	3, 3	92.76	0.000	211.7	7.65	0.000	0.652
Sugar	3, 2	60.91	0.000	90.2	6.75	0.000	0.782
Cocoa	2, 2	81.92	0.000	175.2	7.64	0.000	0.631
Orange Juice	5, 2	37.82	0.000	175.6	5.71	0.000	0.693
Cotton	5, 2	68.17	0.000	215.8	6.75	0.000	0.715
Lumber	1, 2	63.92	0.000	155.6	6.52	0.000	0.608
Financial:							
Deutsche Mark	2, 2	97.44	0.000	379.8	7.23	0.000	0.759
Swiss Franc	2, 3	100.50	0.000	460.7	7.42	0.000	0.769
Japanese Yen	1, 5	73.15	0.000	685.8	6.47	0.000	0.745
British Pound	4, 3	81.07	0.000	466.3	6.52	0.000	0.759
Canadian Dollar	3, 2	59.12	0.000	917.5	6.84	0.000	0.688
Treasury Bills	4, 1	66.43	0.000	219.4	8.15	0.000	0.679
Treasury Bonds	4, 2	106.30	0.000	388.3	8.22	0.000	0.727
Metal/Energy:							
Gold	2, 2	71.74	0.000	282.5	7.59	0.000	0.795
Silver	4, 6	98.77	0.000	201.8	4.71	0.000	0.709
Platinum	2, 2	73.41	0.000	213.4	7.91	0.000	0.703
Heating Oil	1, 1	51.06	0.000	89.4	7.14	0.000	0.645
Crude Oil	4, 1	40.55	0.000	65.5	6.36	0.000	0.683
Gasoline	4, 2	30.15	0.000	119.2	5.03	0.000	0.587

Notes: The model is estimated with OLS, and the Wald χ^2 statistic tests the null, H_0: $b_j = 0, \forall j$. The cumulative impact of returns is calculated, $\Sigma\, b_j$, and tested against the null, H_0: $\Sigma\, b_j = 0$, with a two-tailed t-test. All models are estimated over 536 weekly observations, except for those involving crude oil and gasoline, which are estimated over 438 observations.

pooled cross-sectional time-series models are estimated using the generalized least squares (GLS) procedure of Kmenta (1986, pp. 616–635) correcting for cross-sectional correlation and heteroskedasticity. The lag lengths for the pooled regressions are specified by choosing the maximum p and the maximum q from among the individual market specifications within each group. For instance, in the grain group, the maximum p is 2 (soybeans and soy oil) and the maximum q is 2 (corn, soybeans, soy meal, and soy oil); therefore, the pooled grain model's lag structure is 2, 2. While this specification procedure may overspecify lag structures at the expense of statistical power, it assures the model does not suffer from an underspecification bias.

The estimated pooled models are presented in table 4. For each pooled regression, the null hypothesis that returns do not lead sentiment (i.e., $b_j = 0$, œj) is tested with a Wald Chi-squared test, and the cumulative impact of lagged returns is again tested with a two-tailed t-test (i.e., $\Sigma\, b_j = 0$). Certain important characteristics of sentiment are evident in the results. First, across groups, sentiment follows a fairly strong positive autoregressive process, with first-order coefficients around 0.65. Second, a statistically significant positive relationship between sentiment and returns is demonstrated at one- and two-week lags for all the groups. For instance, in grains, a 1% weekly return results in sentiment increasing by 1.26% the following week and 0.376% the week after that. For all the groups, the null that returns do not lead sentiment can be rejected at the 1% level, and the cumulative impact of lagged returns is significantly positive (1% level).

To illustrate the behavior of sentiment, the impulse response function for a one standard deviation shock to returns is calculated (see Harvey, 1991, p. 234).[7] The graphs in figure 2 show the impulse response functions for the pooled sentiment models. This figure reveals that a one standard deviation shock in weekly returns causes the greatest initial increase in food/fiber market sentiment (panel C).[8] Notably, the impact on metal/energy (panel E) and financial (panel D) market sentiment does not reach a peak until two weeks after the initial shock. All of the response functions decline rather smoothly and at similar rates, except for the livestock group (panel B) where extrapolative effects are less pronounced.[9] In total, the pooled models strongly suggest the sentiment levels are caused by returns, and newsletter writers in aggregate may be trend-followers. These results are consistent with those documented by Solt and Statman (1988) and De Bondt (1993) for retail stock market speculators.

Based on the presented results, two points are clear. First, any test of market sentiment's usefulness as a contrary market indicator must be careful to include only past values of sentiment to avoid the contemporaneous correlations documented in table 1. Second, given the strong evidence that returns lead sentiment, a regression-based predictive model which does not include past returns is potentially misspecified. Therefore, Granger's causality test that sentiment leads returns is a natural extension

---

[7] Implicitly, it is assumed sentiment is endogenous and impacted by an exogenous shock to returns.

[8] The one standard deviation shocks to weekly returns (in parentheses) for each group are as follows: grain (0.029), livestock (0.029), food/fiber (0.042), financial (0.013), and metal/energy (0.036).

[9] The impulse response functions decline toward their long-run or total multiplier which is zero, as is the case for any stationary series.

Table 4. Granger Causality Test for Pooled Futures Market Groups, Returns Lead Sentiment (May 1983–September 1994)

Independent Variable	Market Group				
	Grain	Livestock	Food/Fiber	Financial	Metal/Energy
Intercept	11.09	12.03	10.45	10.33	10.02
	(16.9)	(12.3)	(17.0)	(16.6)	(13.7)
$\rho_{t!1}$	0.664	0.617	0.645	0.692	0.685
	(31.2)	(26.1)	(33.2)	(39.2)	(32.7)
$\rho_{t!2}$	0.091	0.049	0.028	0.021	0.026
	(4.57)	(1.78)	(1.21)	(0.96)	(1.02)
$\rho_{t!3}$	—	0.022	0.044	!0.003	0.029
		(0.80)	(1.94)	(!0.15)	(1.15)
$\rho_{t!4}$	—	0.053	0.011	0.052	0.022
		(2.32)	(0.49)	(3.11)	(1.07)
$\rho_{t!5}$	—	—	0.026	—	—
			(1.54)		
$R_{t!1}$	126.5	95.9	104.0	233.8	94.1
	(14.6)	(11.6)	(18.8)	(17.5)	(13.8)
$R_{t!2}$	37.6	25.9	32.9	79.6	29.5
	(4.44)	(3.03)	(5.64)	(5.67)	(4.15)
$R_{t!3}$	—	4.09	5.22	28.4	4.64
		(0.47)	(0.90)	(2.01)	(0.65)
$R_{t!4}$	—	!5.76	28.50	!1.95	—
		(!0.78)	(2.11)	(!0.28)	
$R_{t!5}$	—	!7.65	!5.93	9.54	—
		(!0.94)	(!0.44)	(1.38)	
$R_{t!6}$	—	!4.67	—	!3.37	—
		(!0.58)		(!0.50)	
Σb_j	164.1	107.8	142.2	364.6	132.4
	(12.80)	(4.86)	(13.20)	(10.70)	(7.24)
$\chi^2_{[q]}$ [a]	221.6	143.9	368.7	364.6	202.7
Buse's R^2	0.667	0.545	0.683	0.671	0.653

Note: The t-statistics (in parentheses) test if the coefficient equals zero, with degrees of freedom equal to $N(K-(p+q+1)$, where $N = 536$ (438 for metal/energy) and K = number of markets in the group.
[a] All the $\chi^2_{[q]}$ statistics reject that the coefficients on lagged returns are zero at the 1% level.

of the work presented thus far, and it is an appropriately specified time-series test for sentiment's value as a contrary indicator.

Tests for Return Predictability

Two different time-series tests of return predictability are employed. The first is a general test based on a Granger causality specification parallel to that presented in equation (1). The second is a more specialized test of contrary opinion based on the market-timing framework developed by Cumby and Modest (1987).

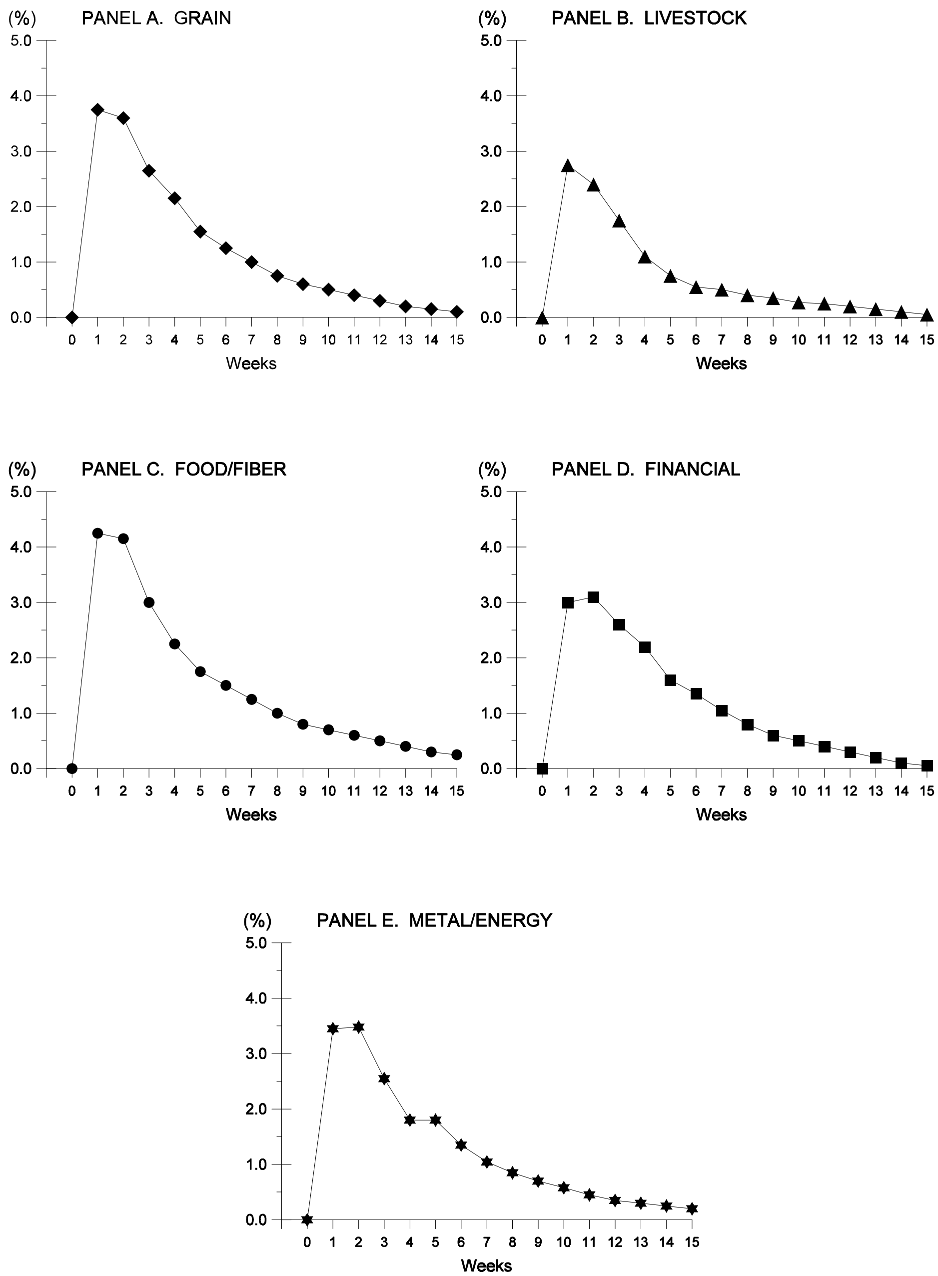

Figure 2. Impulse response functions for futures market groups, return impact on market sentiment (May 1983–September 1994)

Causality Tests

Following the specification and estimation procedures presented for equation (1), the linear linkages between returns and sentiment are examined using the bivariate Granger test:

$$R_t = k_0 + \sum_{i=1}^{m} \alpha_i R_{t-i} + \sum_{j=1}^{n} \beta_j \rho_{t-j} + \varepsilon_t, \tag{2}$$

where R_t and ρ_t are futures returns and noise trader sentiment, respectively, and ε_t is a white noise error term. Sentiment leads returns in equation (2) if market sentiment is useful in predicting returns, and it is tested under the null of $\beta_j = 0, \forall j$. Furthermore, the cumulative impact of market sentiment on returns is tested under the null of $\Sigma\,\beta_j = 0$. Rational expectations is also tested under the full orthogonality condition: $\beta_j = \alpha_i = 0, \forall i, j$. Again, to increase the power of these tests, they are estimated over pooled cross-sectional time-series data using the futures groups presented in table 1.

Equation (2) provides a well-specified and general means of testing the orthogonality condition implied by market rationality. If, however, sentiment is a useful contrary indicator for market returns, then there will be a negative relationship between sentiment and returns—i.e., high (low) sentiment predicts negative (positive) returns. A contrarian relationship should be captured in (2) by finding that sentiment leads returns $\beta_j \neq 0, \forall j$, and the cumulative impact of sentiment on returns is negative ($\Sigma\,\beta_j < 0$). The opposite is true for a positive impact ($\beta_j \neq 0, \forall j$, and $\Sigma\,\beta_j > 0$).

The causality test results for individual markets are presented in table 5. The first χ^2 statistic (column 3) tests the null that sentiment does not lead returns, and the *t*-statistic (column 5) tests if the sum of lagged sentiment coefficients equals zero. The second χ^2 statistic (column 6) tests the full orthogonality condition. The first result of importance is that lagged sentiment did not even enter 14 of the 28 regression models. For the remaining models, the null hypothesis that sentiment does not lead returns is rejected for two markets (lumber and Treasury bills) at the 5% level, and four more markets (feeder cattle, cocoa, orange juice, and live hogs) at the 10% level. The total of six rejections is more than would be expected by chance alone ($0.10 \times 28 = 2.8$ rejections).

While there is some evidence of a relationship between sentiment and subsequent futures returns, the direction of the relationship is not consistent. Trade sources (e.g., *CONSENSUS*) tout sentiment as a contrary indicator, but the *t*-statistics for H_0: $\Sigma\,\beta_j = 0$ reveal that the sum of lagged sentiment coefficients is not consistently negative. The sum is significantly negative in only two cases (live hogs and cocoa). Further, nine of the 14 *t*-statistics are positive, indicating a tendency toward continuation instead of reversal.

The second χ^2 statistic (column 6) in table 5 tests the null hypothesis that neither sentiment nor past returns lead future returns, i.e., returns are not predictable with the information contained in past returns and sentiment. This null is rejected in 13 markets at the 10% level or higher. Of the 13 rejections, eight are in markets where the first χ^2 test did not reject the null, and the rejections are concentrated among the

Table 5. Granger Causality Test for Individual Futures Markets, Sentiment Leads Returns (May 1983–September 1994)

$$\left[\text{Equation (2) causality test: } R_t = k_0 + \sum_{i=1}^{m} \alpha_i R_{t-i} + \sum_{j=1}^{n} \beta_j \rho_{t-j} + \varepsilon_t\right]$$

[1] Market	[2] m, n	[3] $\chi^2_{[n]}$	[4] p-Value	[5] t-Statistic	[6] $\chi^2_{[m+n]}$	[7] p-Value	[8] Adjust. R^2
Grain:							
Corn	5, 0	—	—	—	5.72	0.334	0.017
Wheat	0, 1	0.17	0.679	−0.41	0.17	0.679	−0.001
Soybeans	3, 1	2.30	0.129	−1.51	4.82	0.306	0.008
Soybean Meal	3, 0	—	—	—	4.31	0.230	0.009
Soybean Oil	3, 0	—	—	—	3.45	0.327	0.010
Livestock:							
Live Cattle	6, 0	—	—	—	11.95	0.063	0.015
Feeder Cattle	2, 3	7.48	0.058	0.76	10.61	0.059	0.017
Live Hogs	1, 3	6.39	0.094	−2.05	17.13	0.002	0.024
Pork Bellies	1, 0	—	—	—	0.72	0.393	−0.001
Food/Fiber:							
Coffee	1, 0	—	—	—	0.22	0.638	−0.002
Sugar	0, 1	2.11	0.146	1.45	2.11	0.146	0.002
Cocoa	0, 1	3.21	0.073	−1.79	3.21	0.073	0.004
Orange Juice	1, 5	10.32	0.066	2.62	31.69	0.000	0.041
Cotton	4, 0	—	—	—	12.69	0.012	0.028
Lumber	2, 2	18.68	0.000	−0.49	25.24	0.000	0.059
Financial:							
Deutsche Mark	0, 1	1.21	0.271	1.09	1.21	0.271	0.000
Swiss Franc	3, 0	—	—	—	6.19	0.102	0.009
Japanese Yen	0, 1	2.16	0.141	1.47	2.16	0.141	0.002
British Pound	3, 0	—	—	—	6.86	0.076	0.009
Canadian Dollar	0, 1	0.53	0.462	0.73	0.53	0.462	−0.001
Treasury Bills	0, 5	16.86	0.005	0.06	16.86	0.005	0.015
Treasury Bonds	1, 0	—	—	—	0.64	0.422	−0.001
Metal/Energy:							
Gold	0, 1	0.31	0.574	0.56	0.31	0.574	−0.001
Silver	6, 0	—	—	—	9.32	0.156	0.016
Platinum	6, 1	2.55	0.111	1.59	13.72	0.056	0.015
Heating Oil	3, 0	—	—	—	8.96	0.029	0.022
Crude Oil	3, 0	—	—	—	6.79	0.078	0.013
Gasoline	3, 0	—	—	—	8.77	0.032	0.025

Notes: The model is estimated with OLS, and the first Wald χ^2 statistic (column 3) tests the null, H_0: $\beta_j = 0$, $\forall j$. The t-statistic tests that the sum of the lagged sentiment coefficients equals zero, $\Sigma \beta_j = 0$. The second χ^2 statistic (column 6) tests full orthogonality, H_0: $\alpha_i = 0$ and $\beta_j = 0$, $\forall i, j$. The model is estimated over 536 weekly observations, except for those regressions involving crude oil and gasoline, which have 438 observations.

Table 6. Granger Causality Test for Pooled Futures Market Groups, Sentiment Leads Returns (May 1983–September 1994)

Independent Variable	Market Group, Coefficient × 10[12]				
	Grain	Livestock	Food/Fiber	Financial	Metal/Energy
Intercept	0.0149 (0.10)	!0.0699 (!0.40)	!0.1658 (!0.83)	0.0178 (0.95)	!0.1596 (!0.99)
$R_{t!1}$	0.4028 (0.20)	2.4092 (1.05)	6.5781 (3.34)	!2.7300 (!1.63)	!3.7200 (!1.84)
$R_{t!2}$	5.6255 (2.77)	!2.438 (!1.05)	4.3187 (2.07)	3.8714 (2.29)	3.6352 (1.75)
$R_{t!3}$	7.2097 (3.59)	2.5207 (1.10)	2.4519 (1.17)	2.8223 (1.67)	4.6910 (2.28)
$R_{t!4}$	0.4432 (0.23)	!0.1779 (!0.07)	4.9758 (2.39)	—	1.5412 (0.75)
$R_{t!5}$	!2.1835 (!1.11)	!4.975 (!2.22)	—	—	!4.5359 (!2.25)
$R_{t!6}$	—	2.7824 (1.23)	—	—	!1.8822 (!0.89)
$\rho_{t!1}$	!0.0005 (!0.10)	0.0012 (0.32)	!0.0135 (!2.24)	0.0001 (0.25)	0.0028 (0.89)
$\rho_{t!2}$	—	0.0040 (0.94)	0.0155 (2.16)	!0.0004 (!0.54)	—
$\rho_{t!3}$	—	!0.0003 (!0.09)	!0.0099 (!1.38)	!0.0004 (!0.65)	—
$\rho_{t!4}$	—	—	!0.0010 (!0.14)	0.0013 (2.13)	—
$\rho_{t!5}$	—	—	0.0127 (2.32)	!0.0009 (!1.79)	—
$\chi^2_{[n]}$	0.051 [0.821]	2.76 [0.431]	15.08 [0.010]	5.40 [0.368]	0.80 [0.370]
$\chi^2_{[m\%n]}$	22.28 [0.001]	13.75 [0.131]	35.69 [0.000]	15.26 [0.054]	19.98 [0.005]
Buse's R^2	0.006	0.001	0.009	0.002	0.006

Notes: The *t*-statistics (in parentheses) test if the coefficient equals zero, with degrees of freedom equal to $N(K - (m + n + 1)$, where $N = 536$ (438 for metal/energy) and K = number of markets in the group. The first (second) χ^2 statistic tests H_0: $\beta_j = 0$ (and $\alpha_i = 0$), œ i, j [with *p*-values in brackets].

food/fiber and metal/energy groups. Although not presented, in the markets where the full orthogonality null is rejected, the rejection primarily stems from low-order positive autocorrelation in returns.

The pooled causality results are presented in table 6. Pooled models were estimated with Kmenta's cross-sectionally correlated and heteroskedastic GLS procedure.[10]

[10] The individual models were also estimated in a seemingly unrelated regression (SUR) framework, but the results were not materially different from the OLS estimations.

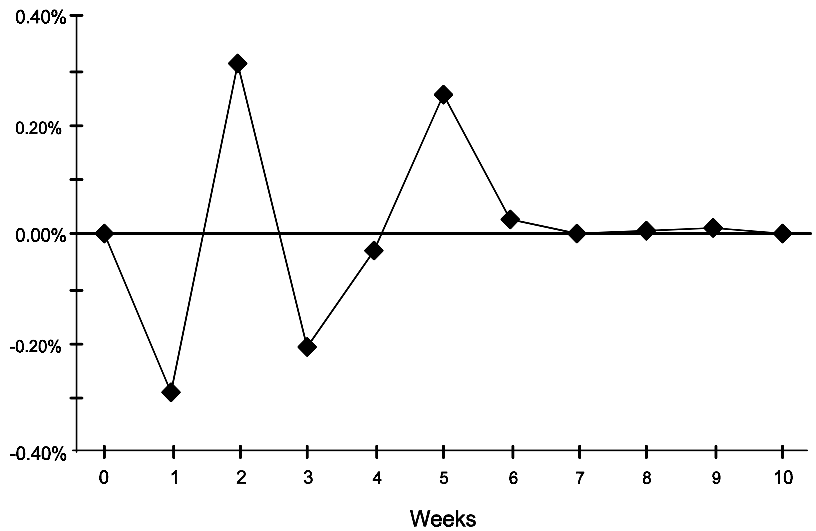

Figure 3. Impulse response function for the food/fiber market group, market sentiment impact on return (May 1983–September 1994)

The first χ^2 statistic (with *p*-values in brackets) tests the null that sentiment does not lead returns, and the second χ^2 statistic tests the full orthogonality condition. The null hypothesis that sentiment does not lead returns is rejected for the food/fiber group only. The full orthogonality null hypothesis is rejected at conventional levels for all groups except livestock. Returns in general, and the food/fiber and grain groups in particular, are characterized by positive autocorrelation at short lags with autoregressive parameters along the order of 0.05 to 0.07 in magnitude.

As with the individual market models, the direction of sentiment's impact on returns is somewhat inconsistent. For example, the food/fiber group's sentiment coefficients are significantly negative at lag one and significantly positive at lags two and five.[11] The full impulse response to a one standard deviation shock to weekly sentiment is plotted in figure 3 for the food/fiber group. The figure shows there is not a well-defined response structure for sentiment leading returns. That is, the response function takes both positive and negative values before converging to zero after seven weeks.

Collectively, the causality models provide some mild evidence that newsletter sentiment is useful in predicting market returns. However, the null hypothesis is rejected in a relatively small number of the markets. Furthermore, the direction of sentiment's impact is not consistent across markets. The small amount of evidence which does exist would suggest price continuation over weekly intervals, not price reversals. This evidence is not supportive of using sentiment as a contrary indicator.

[11] The sum of lagged sentiment coefficients is not significantly different from zero for the food/fiber group.

However, the findings are consistent with those reported by Wang (2001) for CFTC small speculators in agricultural futures markets.

Cumby-Modest Test

While the causality test results presented above do not indicate a consistent relationship between noise trader sentiment and subsequent futures price movements, it may be possible that a relationship exists, but only at extreme levels of sentiment (*CONSENSUS*; Wang, 2001). The market-timing framework proposed by Cumby and Modest (1987) can be used to determine whether extreme sentiment readings provide market signals. Given a definition of extremely high and low sentiment levels (K_H and K_L, respectively), the Cumby-Modest (C-M) test is based on the following OLS regression:

$$(3) \qquad R_t = \alpha + \beta_1 HI_{t-1} + \beta_2 LO_{t-1} + \varepsilon_t ,$$

where $HI_{t-1} = 1$ if $\rho_{t-1} > K_H$, and $HI_{t-1} = 0$ otherwise; $LO_{t-1} = 1$ if $\rho_{t-1} < K_L$, and $LO_{t-1} = 0$ otherwise. If the mean return conditioned on extreme optimism ($\alpha + \beta_1$) or pessimism ($\alpha + \beta_2$) is different from the unconditional mean (α), then timing ability is demonstrated. The null hypothesis of no predictability, H_0: $\beta_1 = \beta_2 = 0$, is tested against the alternative of significant timing ability, H_A: $\beta_1 \neq 0$ or $\beta_2 \neq 0$. "Contrary opinion" would suggest that $\beta_1 < 0$ or $\beta_2 > 0$, indicating extreme sentiment is negatively related to returns.

Consensus, Inc. suggests that sentiment outside the range of (25, 75) denotes a market approaching extreme conditions. For the initial C-M tests, extreme sentiment is defined by these levels plus a factor of five to assure that the extremes compose a small percentage of the total observations. The C-M test results for individual markets with $K_H = 80$ and $K_L = 20$ are presented in table 7.[12] For individual markets, the number of extreme observations constitutes from 4.3% (23) to 30% (161) of the 536 total observations for each market. Based on χ^2 statistics, the null hypothesis of no timing ability ($\beta_1 = \beta_2 = 0$) is rejected for three markets (live cattle, Canadian dollar, gasoline) at the 5% level and two more markets (soybeans, cocoa) at the 10% level. Again, the five rejections are more than would be expected by chance ($0.10 \times 28 = 2.8$ rejections). There also are four cases where an individual coefficient is significantly different from zero (β_2 for wheat, Japanese yen, platinum, and crude oil), but the joint test is insignificant. Finally, it is worth noting that only one of the rejections (cocoa) is common to both the C-M and causality tests.

While there is evidence of a significant relationship between extreme sentiment and returns, the direction of the relationship is, if anything, one of continuation.

---

[12] The OLS error terms are tested for heteroskedasticity using White's test, and for autocorrelation using the Lagrange multiplier test. If the errors are heteroskedastic, then the model is estimated using White's heteroskedastic consistent covariance estimator, and if the errors are autocorrelated, then the Newey-West covariance estimator is utilized (Hamilton, 1994, p. 281).

Table 7. Cumby-Modest Test for Individual Futures Markets (May 1983–September 1994)

[Equation (3) Cumby-Modest regression: R_t ' α % $\beta_1 HI_{t\&1}$ % $\beta_2 LO_{t\&1}$ % g_t]

Market	Extreme Observations	α × 10!2	β_1 × 10!2	β_2 × 10!2	$\chi^2_{[2]}$	*p*-Value
Grain:						
Corn	76	!0.1472	0.7522	!0.0672	1.09	0.579
		(!1.09)	(1.03)	(!0.18)		
Wheat	92	!0.0869	0.4283	0.6477	4.05	0.131
		(!0.71)	(0.82)	(1.89)		
Soybeans	47	!0.1400	0.5031	0.7299	5.36	0.068
		(!1.03)	(0.49)	(2.27)		
Soybean Meal	106	0.0198	!0.3885	!0.1124	0.42	0.801
		(0.12)	(!0.55)	(!0.38)		
Soybean Oil	126	0.0538	0.5080	!0.5283	2.99	0.223
		(0.29)	(0.50)	(!1.60)		
Livestock:						
Live Cattle	23	0.1659	!0.0224	2.0730	9.83	0.007
		(1.88)	(!0.06)	(3.13)		
Feeder Cattle	78	0.0831	0.1825	0.2078	0.72	0.696
		(1.02)	(0.71)	(0.56)		
Live Hogs	23	0.2596	!0.9115	!0.0704	0.39	0.822
		(2.02)	(!0.61)	(!0.10)		
Pork Bellies	88	!0.3557	1.3560	0.1414	0.74	0.689
		(!1.49)	(0.84)	(0.22)		
Food/Fiber:						
Coffee	113	!0.2414	0.3544	0.1850	0.21	0.901
		(!1.24)	(0.23)	(0.40)		
Sugar	117	!0.4081	0.8003	!0.9866	2.54	0.279
		(!1.36)	(0.87)	(!1.21)		
Cocoa	110	!0.4495	0.9907	0.9082	4.82	0.089
		(!2.36)	(0.91)	(2.05)		
Orange Juice	161	0.2338	0.5623	!0.5112	3.73	0.154
		(1.18)	(0.83)	(!1.63)		
Cotton	101	0.1648	0.5484	!0.4501	3.54	0.170
		(1.13)	(0.99)	(!1.46)		
Lumber	120	0.0235	!1.2387	!0.0762	1.18	0.553
		(0.12)	(!1.08)	(!0.18)		
Financial:						
Deutsche Mark	105	0.0626	0.7518	!0.7747	0.23	0.890
		(0.74)	(0.21)	(!0.40)		
Swiss Franc	115	0.0103	0.2532	0.0434	0.56	0.756
		(0.11)	(0.73)	(0.21)		
Japanese Yen	98	0.1735	!0.2310	!0.3250	3.22	0.198
		(2.38)	(!0.70)	(!1.71)		

(continued . . .)

Table 7. Continued

Market	Extreme Observations	$\alpha \times 10^{-2}$	$\beta_1 \times 10^{-2}$	$\beta_2 \times 10^{-2}$	$\chi^2_{[2]}$	p-Value
Financial (cont'd):						
British Pound	130	0.0355	0.3856	−0.0264	2.23	0.328
		(0.39)	(1.46)	(−0.11)		
Canadian Dollar	98	0.0582	−0.1064	−0.1877	6.29	0.043
		(2.13)	(−0.72)	(−2.46)		
Treasury Bills	98	0.0019	0.0305	−0.0190	2.17	0.337
		(2.21)	(1.34)	(−0.48)		
Treasury Bonds	39	0.0173	−0.5901	−0.4312	2.09	0.351
		(2.44)	(−0.11)	(−1.44)		
Metal/Energy:						
Gold	101	−0.0180	0.4368	0.0685	0.75	0.687
		(−2.00)	(0.84)	(0.24)		
Silver	68	−0.4169	0.7366	0.5552	1.58	0.452
		(−2.81)	(0.80)	(0.96)		
Platinum	114	−0.4510	0.3082	−0.5878	3.26	0.196
		(−0.32)	(0.29)	(−1.76)		
Heating Oil	130	0.0923	−1.1126	−0.6281	1.18	0.553
		(0.43)	(−1.08)	(−0.12)		
Crude Oil	67	0.3340	−0.8007	−1.7014	3.36	0.186
		(1.38)	(−0.94)	(−1.67)		
Gasoline	120	0.4406	−0.1653	−0.8561	6.87	0.032
		(1.82)	(−2.44)	(−1.33)		

Notes: The model is estimated with OLS, where $HI_{t-1} = 1$ if $\rho_{t-1} > K_H$, and $HI_{t-1} = 0$ otherwise; $LO_{t-1} = 1$ if $\rho_{t-1} < K_L$, and $LO_{t-1} = 0$ otherwise; and $K_H = 80$, $K_L = 20$. Values in parentheses are t-statistics, and the Chi-squared test is a joint test of the null, H_0: $\beta_1 = \beta_2 = 0$. All models are estimated over 536 weekly observations, except for those involving crude oil and gasoline, which are estimated over 438 observations.

Returns increase (decrease) after high (low) sentiment, rather than reverse. In addition, there is variation in the coefficient signs for those markets where the null is rejected. For instance, if the CBSI is below 20, then the following week nearby live cattle returns increase by 2.07% on average, while Canadian dollar returns fall 0.188%.[13]

Pooled C-M test results with $K_H = 80$ and $K_L = 20$ are presented in panel B of table 8. The pooled C-M models are estimated using Kmenta's cross-sectionally correlated, heteroskedastic, and timewise autoregressive GLS estimation technique. Of the five market groups, the null of no market timing is rejected at the 5% level for the grains and at the 10% level for the financials. For example, the estimated coefficients show weekly grain futures returns increase by 0.4029% after sentiment

[13] In the text, the C-M coefficients are always referred to as the change in returns or expected percentage price change, relative to the unconditional return. This is in contrast to the total expected return. For instance, when the CBSI is below 20, the expected weekly live cattle return increases by 2.07%, but the total expected return is 2.24% (2.07 + 0.17).

Table 8. Cumby-Modest Test for Pooled Futures Market Groups (May 1983–September 1994)

[Equation (3) Cumby-Modest regression: $R_t = \alpha + \beta_1 HI_{t-1} + \beta_2 LO_{t-1} + \varepsilon_t$]

PANEL A. $K_H = 75$, $K_L = 25$

Group	$\alpha \times 10^{-2}$	$\beta_1 \times 10^{-2}$	$\beta_2 \times 10^{-2}$	$\chi^2_{[2]}$	*p*-Value
Grain	−0.0416	0.3687	0.0273	8.16	0.016
	(−0.43)	(2.83)	(0.36)		
Livestock	0.1239	0.2009	0.1462	3.69	0.157
	(1.57)	(1.53)	(1.20)		
Food/Fiber	−0.0245	−0.0703	0.0239	0.16	0.943
	(−0.28)	(−0.28)	(0.15)		
Financial	0.0096	0.0087	0.0089	0.36	0.834
	(1.16)	(0.37)	(0.51)		
Metal/Energy	−0.0263	0.1259	−0.1833	3.78	0.151
	(−0.29)	(0.74)	(−1.75)		

PANEL B. $K_H = 80$, $K_L = 20$

Group	$\alpha \times 10^{-2}$	$\beta_1 \times 10^{-2}$	$\beta_2 \times 10^{-2}$	$\chi^2_{[2]}$	*p*-Value
Grain	−0.0308	0.4029	0.0689	6.44	0.039
	(−0.32)	(2.43)	(0.73)		
Livestock	0.1519	−0.0070	−0.0338	0.04	0.978
	(1.95)	(−0.04)	(−0.20)		
Food/Fiber	−0.0463	0.4979	0.0116	2.36	0.307
	(−0.58)	(1.54)	(0.06)		
Financial	0.0117	0.0549	−0.0209	4.65	0.098
	(1.52)	(1.82)	(−1.02)		
Metal/Energy	−0.0362	0.2118	−0.2311	4.07	0.131
	(−0.41)	(0.86)	(−1.80)		

PANEL C. $K_H = 85$, $K_L = 15$

Group	$\alpha \times 10^{-2}$	$\beta_1 \times 10^{-2}$	$\beta_2 \times 10^{-2}$	$\chi^2_{[2]}$	*p*-Value
Grain	−0.0113	0.4166	0.0027	2.49	0.286
	(−0.12)	(1.58)	(0.02)		
Livestock	0.1563	−0.0598	−0.2856	1.72	0.423
	(2.02)	(−0.22)	(−1.29)		
Food/Fiber	−0.0362	−0.0515	0.1294	0.36	0.833
	(−0.46)	(−0.11)	(0.58)		
Financial	0.0122	0.0561	−0.0156	1.91	0.385
	(1.60)	(1.19)	(−0.65)		
Metal/Energy	−0.0393	0.1591	−0.2148	1.98	0.372
	(−0.45)	(0.47)	(−1.32)		

Notes: The model is estimated over N cross-sections and T time-series observations, where $HI_{t-1} = 1$ if $\rho_{t-1} > K_H$, and $HI_{t-1} = 0$ otherwise; $LO_{t-1} = 1$ if $\rho_{t-1} < K_L$, and $LO_{t-1} = 0$ otherwise. The *t*-statistics (in parentheses) test that parameter values are zero, and the χ^2 tests the joint null, H_0: $\beta_1 = \beta_2 = 0$. All models are estimated over 536 weekly observations, except the metal/energy group, which is estimated over 438 observations. Each pooled regression has $N \times T$ cross-sectional time-series observations, where $T = 536$ (or 438) and N is the number of markets comprising the group.

readings above 80%, and increase by 0.0689% after sentiment readings below 20%.[14]

Parameter sensitivity is explored by altering the definition of extreme sentiment. In panel A of table 8, $K_H = 75$ and $K_L = 25$, while in panel C, $K_H = 85$ and $K_L = 15$. At decreased extremes (panel A), the grain model still displays statistically significant timing ability, but when the extreme sentiment definitions are widened (panel C), none of the pooled models reject the null hypothesis. These results suggest that the impact of market sentiment is quite sensitive to alternative definitions of extreme index values.

Overall, the C-M test results indicate some evidence of a predictive relationship between extreme sentiment and subsequent futures returns. The evidence is strongest for grain futures markets. However, the relationship is indicated for only a limited number of the markets, it is sensitive to the definition of "extreme" sentiment, and the direction of extreme sentiment's impact generally is that of continuation. These results do not support the use of the sentiment index as a contrary market indicator, even at extreme levels. The tenuous causal relationships between sentiment and returns are consistent with Fisher and Statman's (2000) results for newsletter writers in equity markets and Wang's (2001) findings for CFTC small speculators in agricultural futures.

Summary and Conclusions

The theory of contrary opinion predicts price reversals following extremes in market sentiment. This analysis tests return predictability in futures markets using a direct, or survey-based, measure of sentiment. Findings show the Consensus Index of Bullish Market Opinion primarily reflects the opinions of newsletter writers and the corresponding market positions of their primary audience—retail futures speculators. The lead-lag relationships between sentiment and futures market returns are investigated within a Granger causality framework. The results suggest that sentiment is an increasing function of past futures returns over at least the previous two weeks, and retail futures speculators may be trend-followers who act in unison to correlated market signals (past returns). This characterization is consistent with the theoretical "noise traders" of De Long et al. (1990a, b), and the results can have ramifications for interpreting and testing noise trader models (Sanders, Irwin, and Leuthold, 2000).

In this research, market sentiment is found to have little predictive power in futures markets. Time-series regressions (Granger causality tests) are specified to test if the level of noise trader sentiment consistently predicts subsequent futures price movements. The predictive ability of extremely high or low sentiment is also tested in a Cumby-Modest (1987) market-timing framework. The time-series regressions provide some evidence that noise trader sentiment is useful in predicting market returns,

[14] While these returns are statistically significant, their economic significance is debatable. It is not the intent of this study to search for an economically significant trading strategy. Instead, we are simply investigating the time-series predictability of returns.

particularly when sentiment is at extreme levels. However, relationships are present for only a limited number of the markets, and the direction of sentiment's impact generally is inconsistent. The relationship, if any, tends toward continuation. Specifically, there is little or no evidence supporting market reversals or "contrary opinion." This conclusion is fairly consistent across the 28 futures markets examined. Therefore, practitioners who rely solely on this type of indicator may be misguided. The results may also provide interpretive evidence for different theoretical models which suggest sentiment can influence market returns (e.g., Cutler, Poterba, and Summers, 1989; Barberis, Shleifer, and Vishny, 1998).

This research focuses on one sentiment index, the Consensus Index of Bullish Market Opinion, which reflects the sentiment of market newsletters. As pointed out by Fisher and Statman (2000), the results could vary using the sentiment of another segment of market participants (see also Wang, 2001). Likewise, sentiment could impact other aspects of price behavior, such as volatility. We concur with Fisher and Statman that this line of research needs to be broadened to additional markets, alternative forms of price behavior, and other measures of sentiment.

References

Barberis, N., A. Shleifer, and R. Vishny. (1998). "A model of investor sentiment." *Journal of Financial Economics* 49, 307–343.

Beveridge, S., and C. Oickle. (1994). "A comparison of Box-Jenkins and objective methods for determining the order of a non-seasonal ARMA Model." *Journal of Forecasting* 13, 419–434.

Brennan, M. J. (1995). "The individual investor." *Journal of Financial Research* 18, 59–74.

Brown, G. W. (1999). "Volatility, sentiment, and noise traders." *Financial Analysts Journal* 55, 82–90.

Brown, G. W., and M. T. Cliff. (2001, September). "Investor sentiment and the near-term stock market." Financial Economics Network, Capital Markets Abstracts: Market Efficiency. Working paper, Kenan-Flagler Business School, University of North Carolina-Chapel Hill.

Canoles, B. W., S. Thompson, S. Irwin, and V. G. France. (1998). "An analysis of the profiles and motivations of habitual commodity speculators." *Journal of Futures Markets* 18, 765–801.

Clarke, R. G., and M. Statman. (1998). "Bullish or bearish." *Financial Analysts Journal* 54, 63–72.

Consensus, Inc. (1995, February 17). *CONSENSUS: National Futures and Financial Weekly* [Kansas City, MO], Vol. XXV, No. 7.

Consensus, Inc. (2001, November 15). Sentiment index. Online website. Available at http://www.consensus-inc.com/.

Cumby, R. E., and D. M. Modest. (1987). "Testing for market timing ability: A framework for forecast evaluation." *Journal of Financial Economics* 19, 169–189.

Cutler, D. M., J. M. Poterba, and L. H. Summers. (1989). "Speculative dynamics and the role of feedback traders." *American Economic Review* 80, 63–68.

Daily % Bullish. (2000, October). Online website. Available at http://www.homestead.com/dailybullish/.

De Bondt, W. F. M. (1993). "Betting on trends: Intuitive forecasts of financial risk and return." *International Journal of Forecasting* 9, 355–371.

De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann. (1990a). "Noise trader risk in financial markets." *Journal of Political Economy* 98, 703–738.

———. (1990b). "Positive feedback investment strategies and destabilizing rational speculation." *Journal of Finance* 45, 379–395.

Draper, D. W. (1985). "The small public trader in futures markets." In A. E. Peck (ed.), *Futures Markets: Regulatory Issues* (pp. 211–269). Washington, DC: American Enterprise Institute for Public Policy Research.

Elton, E. J., M. J. Gruber, and J. A. Busse. (1998). "Do investors care about sentiment?" *Journal of Business* 71, 477–501.

Fisher, K. L., and M. Statman. (2000). "Investor sentiment and stock returns." *Financial Analysts Journal* 56, 16–23.

Greene, W. H. (1993). *Econometric Analysis.* New York: MacMillan Publishing Co.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton, NJ: Princeton University Press.

Harvey, A. C. (1991). *The Econometric Analysis of Time Series*, 2nd edition. Cambridge, MA: MIT Press.

Jones, J. (1989). "A comparison of lag-length selection techniques in tests of Granger causality between money growth and inflation: Evidence for the U.S., 1959–86." *Applied Economics* 24, 809–822.

Kmenta, J. (1986). *Elements of Econometrics*, 2nd edition. New York: Macmillan Publishing Co.

Nagy, R. A., and R. W. Obenberger. (1994, July/August). "Factors influencing individual investor behavior." *Financial Analysts Journal* 50, 63–68.

Neal, R., and S. M. Wheatley. (1998). "Do measures of investor sentiment predict returns?" *Journal of Financial and Quantitative Analysis* 33, 523–547.

Neill, H. (1960). *The Art of Contrary Thinking.* Caldwell, OH: Caxton Printers.

Sanders, D. R., S. H. Irwin, and R. M. Leuthold. (2000). "Noise trader sentiment in futures markets." In B. A. Goss (ed.), *Models of Futures Markets* (pp. 86–116). New York: Routledge.

Simon, D. P., and R. A. Wiggins. (2001). "S&P futures returns and contrary sentiment indicators." *Journal of Futures Markets* 21, 447–462.

Smidt, S. (1965). "Amateur speculators." Cornell Studies in Policy Administration, Cornell University, Ithaca, NY.

Solt, M. E., and M. Statman. (1988, September). "How useful is the sentiment index?" *Financial Analysts Journal* 44, 45–55.

Teweles, R. J., and F. J. Jones. (1999). *The Futures Game: Who Wins, Who Loses, and Why.* New York: McGraw-Hill.

Thorton, D., and D. Batten. (1985). "Lag-length selection and tests of Granger causality between money and income." *Journal of Banking and Finance* 17, 164–178.

Wang, C. (2001). "Investor sentiment and return predictability in agricultural futures markets." *Journal of Futures Markets* 21, 929–952.