Journal of Banking and Finance 90 (2018) 17–31

Contents lists available at ScienceDirect

Journal of Banking and Finance

journal homepage: www.elsevier.com/locate/jbf

Investor sentiment and price discovery: Evidence from the pricing dynamics between the futures and spot markets☆

Chu-Bin Lin[a,*], Robin K. Chou[b], George H.K. Wang[c]

[a] Department of Finance, School of Economics and Management, Southwest Jiaotong University, 111, Section 1, North Erhuan Road, Chengdu 610031, China
[b] Department of Finance and Risk and Insurance Research Center, College of Commerce, National Chengchi University, Taiwan
[c] Department of Finance, School of Business, George Mason University, United States

ARTICLE INFO

Article history:
Received 2 May 2017
Accepted 23 February 2018

JEL classification:
G02
G12
G14

Keywords:
Information shares
Investor sentiment
Lead–lag relation
Limits to arbitrage
Price discovery

ABSTRACT

This study examines the role of investor sentiment in the pricing dynamics between the spot and futures markets. The empirical evidence suggests that investor sentiment has a positive impact on price volatility and the bid–ask spread on both the spot and futures markets, which induces higher arbitrage risk and trading costs during high sentiment periods. Consequently, during high sentiment periods, informed traders become less willing to leverage their information advantages on the futures market, which diminishes the futures markets' leading informational role and contributions to price discovery. Our findings provide support for the theory of limits to arbitrage.

1. Introduction

This paper investigates the relation between investor sentiment and the price discovery process between the spot and futures markets. Extensive literature finds that futures prices lead spot prices and contribute more to the price discovery process (Chan, 1992; Frino et al., 2000). However, far less clear is how this leading role of futures markets in price discovery will be affected as trading risks and/or costs fluctuate over time. This paper aims to fill this gap in the literature by characterizing the relation between investor sentiment, trading risks and costs, and the price discovery process, and we propose that investor sentiment can significantly affect cross-market pricing dynamics.

When homogeneous or closely linked securities trade in multiple markets, price discovery tends to occur first in the market where informed traders can utilize their information with the highest net profit. The literature shows that the futures market, compared to the spot market, offers higher leverages, lower trading costs, and fewer short-sale restrictions (Black, 1975; Kawaller et al., 1987; Stoll and Whaley, 1990; Back, 1993; Mayhew et al., 1995; Fleming et al., 1996; Käppi, 1997; Easley et al., 1998). Accordingly, informed traders prefer the futures market, and the futures market reacts to new information faster than the spot market does. Several studies find that futures returns significantly lead spot index returns in the price discovery process (Finnerty and Park, 1987; Ng, 1987; Kawaller et al., 1987; Harris, 1989; Stoll and Whaley, 1990; Chan, 1992).

On the other hand, the literature shows that investor sentiment affects investors' trading behavior (Tetlock, 2007; Kurov, 2008; Garcia, 2013) and has a significant impact on stock returns and price volatility (Lee et al., 2002; Brown and Cliff, 2005; Baker and Wurgler, 2006; Schmeling, 2009; Kurov, 2010; Berger and Turtle, 2011; Baker et al., 2012; Garcia, 2013). Yu and Yuan (2011) and Stambaugh et al. (2012) suggest that high investor sentiment at-

☆ The authors thank Jarrad Harford, Michael Hemler, Hung-Chia Hsu, Chuan-Yang Hwang, Ji-Chai Lin, Alexander Ljungqvist, Andy Puckett, Michael Sikorski, Giorgio Valente, and seminar participants at the 2014 Midwest Finance Association meeting in Orlando, the 2014 Research in Behavioural Finance Conference in Rotterdam, the 2014 SFM Conference in Kaohsiung, the 2015 FMA European Conference in Venice, and the 2015 FMA annual meeting in Orlando for their helpful comments and suggestions. Chu-Bin Lin gratefully acknowledges the financial support from the Ministry of Science and Technology of Taiwan (Nos. 101-2420-H-004-028-DR and 102-2917-I-004-002) and from the National Natural Science Foundation of China (Nos. 91746109 and 71773100). Robin K. Chou gratefully acknowledges the financial support from the Ministry of Science and Technology of Taiwan (No. 101-2410-H-004-067-MY3) and from the National Natural Science Foundation of China (Nos. 71232004, 71373296, 71372137, and 70902030).

* Corresponding author.
E-mail address: cblin@swjtu.edu.cn (C.-B. Lin).

https://doi.org/10.1016/j.jbankfin.2018.02.014
0378-4266/

tracts more noise traders into the market, which increases noise trader risk and undermines market efficiency (Barberis et al., 1998; Brown, 1999; Karlsson et al., 2009; Yuan, 2015).

Noise trader risk, driven by high investor sentiment, plays an important role in determining the participation of informed traders. Specifically, the theory of limits to arbitrage predicts that informed traders are less willing to utilize their information when noise trader risk is high (Shleifer and Vishny, 1997). De Long et al. (1990) show that the unpredictability of noise traders' future beliefs creates a risk in asset prices that deters rational arbitrageurs from taking aggressive positions against them. Similarly, Barberis et al. (1998) argue that informed traders who bet against mispricing run the risk that investor sentiment will become more extreme and that prices will move even further away from fundamental values. Thus, noise trader risk can be a source of limits to arbitrage, preventing informed traders from utilizing their information.

Given that high investor sentiment increases noise trading and noise trader risk, which are positively related to trading risks and costs, we posit that informed traders become less willing to leverage their information on the futures market during high sentiment periods. This phenomenon, in turn, diminishes the futures market's leading informational role (in the short run) and contribution to price discovery (in the long run) during high sentiment periods.

We begin our empirical analyses by validating investor sentiment as a positive shock to both trading risks and costs. First, we use price volatility as a proxy for trading risks (Brown, 1999; Lee et al., 2002; Yuan, 2015). Jones and Seguin (1997) find that because noise trades are not based on information about underlying values, these trades can move prices away from their intrinsic value, thus increasing price volatility.

Next, we use bid–ask spread as a proxy for trading costs. The literature suggests three major determinants of bid–ask spreads: order processing costs, adverse selection costs, and inventory risk costs. The relation between investor sentiment and the bid–ask spread is somewhat complicated and ambiguous because investor sentiment can affect the bid–ask spread through two possible channels. On one hand, increased noise trading driven by high investor sentiment lowers adverse selection costs, which, in turn, decreases the bid–ask spread (Glosten and Milgrom, 1985; Lee et al., 1993; Greene and Smart, 1999; Baker and Stein, 2004; Liu, 2015). The literature suggests that when liquidity providers encounter a pool of noise traders during high sentiment periods, they decrease the bid–ask spread, resulting a negative impact of investor sentiment on the bid–ask spread.

On the other hand, investor sentiment can positively impact the bid–ask spread due to the positive relation between price volatility and the inventory risk component. Stoll (1978) proposes that inventory costs comprise the price risk and opportunity costs of holding securities. Both theoretical and empirical studies suggest a positive relation between price volatility and the bid–ask spread (Amihud and Mendelson, 1987; Roll, 1984; French and Roll, 1986; Glosten, 1987). Moreover, Cohen et al. (1986) review previous empirical studies and find, in general, a positive relation between transaction price volatility and bid–ask spreads due to the relative importance of order processing costs and the inventory risk component.[1] Given that investor sentiment positively affects price volatility and the positive relation between price volatility and the inventory risk component, we expect a positive impact of investor sentiment on the bid–ask spread.

Intuitively, the net effect of investor sentiment on the bid–ask spread is an empirical issue: The relation between investor sentiment and the bid–ask spread is negative (positive) if the adverse selection (inventory risk) effect dominates the inventory risk (adverse selection) effect. However, Glosten and Milgrom (1985) show that the adverse selection effect exists only when liquidity providers in the market can identify potential noise traders. Whereas prior studies find that, in some specific information events, liquidity providers can easily identify shifts in trader components and adjust the bid–ask spread accordingly, Greene and Smart (1999) suggest that large shifts in noise and informed trading are difficult to identify. Given that our empirical examination does not focus on any specific information event and that investor sentiment is unobservable, we expect the inventory risk effect to dominate the adverse selection effect, resulting in a positive relation between investor sentiment and the bid–ask spread.

To examine the impact of investor sentiment on the pricing dynamics between the spot and futures markets, we use Standard & Poor's Depositary Receipts (S&P 500 ETFs), the Nasdaq 100 Index Tracking Stocks (Nasdaq 100 ETFs), and the unit investment trust of the Dow Jones Industrial Average (DJIA ETFs) along with their corresponding futures contracts. Our analysis leads to several findings. First, investor sentiment has a positive impact on both price volatility and the bid–ask spread. The minute-by-minute realized volatility of exchange-traded funds (ETFs) and futures contracts is significantly positively related to investor sentiment. This finding is consistent with prior research showing that sentiment-driven investors trade more actively during high sentiment periods and thereby increase noise in the market (Jones and Seguin, 1997; Barberis et al., 1998; Brown, 1999; Karlsson et al., 2009; Yuan, 2015). Moreover, the bid–ask spreads of ETFs and futures contracts are, in general, significantly positively related to investor sentiment, implying that the inventory risk effect dominates the adverse selection effect and that trading costs increase during high sentiment periods.

Second, in the short run, the leading informational role of futures is significantly weaker when investor sentiment is high, consistent with the argument that informed investors tend to trade less aggressively on the futures markets when trading risks and costs increase. This effect is not only statistically significant but also economically significant. In one of our empirical tests for the lead–lag relation between the DJIA ETFs and futures, the coefficient on the first lagged futures return drops by 23%, with the ETFs returns as the dependent variable, when the sentiment index is higher than the 75th percentile in our sample period.

Third, in the long run, futures market information shares are negatively related to investor sentiment. This finding suggests that investor sentiment not only has temporal effects on the lead–lag relation but also affects the price discovery process between the two markets in the long-run equilibrium. Prior literature shows that futures prices contribute more to price discovery in equilibrium than spot prices do, suggesting that futures prices are more informative (Chan, 1992; Frino et al., 2000). Our evidence, however, shows that futures prices become relatively less informative when investor sentiment is high, which is in line with De Long et al. (1990), Chan (1992), Fleming et al. (1996), and Barberis et al. (1998), who argue that informed traders tend to be less willing to utilize their information when noise trader risk is high.

Our results are robust when we use alternative sentiment index. Overall, the findings are in line with the theory of limits to arbitrage. While Kyle's (1985) standard model suggests that a higher level of noise trading may allow informed traders to extract more value from their information, our empirical evidence shows that informed traders are less willing to utilize their information on the futures market during high sentiment periods due to increased trading risk and costs. Consequently, during high sentiment periods, the leading informational role of the futures market is diminished, and the futures market contributes relatively less to the

[1] Huang and Stoll (1997) empirically estimate the relative importance of these three components: adverse selection component (9.59%), order processing component (61.76%) and inventory risk component (28.65%).

price discovery process. Chan (1992) suggests that the lead–lag relation between the futures and spot markets may change over time as the costs incurred by informed traders to exploit their information on the market fluctuate. This study adds to the literature by empirically demonstrating that the time-varying lead–lag relation between the futures and spot markets is due, at least in part, to the impact of time-varying investor sentiment and by showing that investor sentiment affects not only price volatility and bid–ask spread within a single market but also information transmission across closely linked markets.

The remainder of this paper is organized as follows. Section 2 reviews related literature on the lead–lag relation between the spot and futures markets, investor sentiment, and the theory of limits to arbitrage and proposes our hypotheses. Section 3 describes the data and measurement of variables used in this study, and Section 4 discusses our methodology. Section 5 presents the empirical results. Section 6 concludes.

2. Literature review and hypothesis

2.1. The lead–lag relation

In perfectly frictionless and rational markets, the price movements of securities and their derivatives simultaneously reflect new information. However, because real-world friction exists and trading costs differ across markets, security and derivative prices can differ and trading activity can vary across markets. Fleming et al. (1996) propose a trading cost hypothesis and suggest that, given the differences in trading costs, price discovery will tend to occur first in the lowest-cost market, as information-based trades are executed in markets with the highest potential net profit.

As mentioned in the previous section, due to the advantages of facilitating informed trading in the futures market, the literature finds that the futures market reacts more quickly to new information. Empirical evidence shows that futures returns significantly lead spot returns, yet spot returns only have limited predictability on futures returns. Kawaller et al. (1987) find that the S&P 500 futures lead the spot index by between 20 and 45 minutes whereas little evidence suggests that the spot index leads the futures index. Also, Stoll and Whaley (1990) find that the S&P 500 and the Major Market Index futures lead the stock indexes by about 5 minutes whereas the feedback time from the spot market to the futures market is much shorter. In general, the literature suggests that the asymmetric lead–lag relation is due to informed traders' preference to trade on the futures market.[2]

2.2. Investor sentiment, noise trader risk, and limits to arbitrage

Empirical evidence shows that investor sentiment plays an important role in financial markets and can predict asset returns both in the spot and futures markets. In the spot market, Brown and Cliff (2005), Baker and Wurgler (2006), and Tetlock (2007) show that investor sentiment can predict stock returns and/or trading volume, indicating that investor sentiment is closely related to stock mispricing. Garcia (2013) further shows that the effect of sentiment on predictability of stock returns is mostly concentrated in recessions, which implies that the effect of investor sentiment is state-contingent. In the futures market, Smales (2014) finds that sentiment has a significant impact on returns in gold futures market with negative sentiment invoking a greater contemporaneous response in returns.

The literature suggests that investor sentiment can predict asset returns because it is related to stock mispricing and noise trading. Barberis et al. (1998) argue that noise traders increase their participation in the market when investor sentiment is high. Karlsson et al. (2009) and Yuan (2015) find evidence that sentiment-driven investors trade more aggressively during high sentiment periods. Following this strand of literature, Yu and Yuan (2011) show that the mean–variance tradeoff becomes much weaker during high sentiment periods because sentiment-driven traders are inexperienced and more likely to misestimate the variance of returns, which weakens the mean–variance tradeoff accordingly. Stambaugh et al. (2012) examine the profitability of long–short strategies on 11 market anomalies (e.g., failure probability, net stock issues, total accruals, momentum, asset growth, return on assets, etc.) and find that each anomaly is stronger (i.e., the long–short strategy is more profitable) following high levels of sentiment. Yu and Yuan and Stambaugh et al. both suggest that the market is less efficient during high sentiment periods due to the higher participation of noise traders.

Noise trader risk driven by noise trading during high sentiment periods is likely a source of limits to arbitrage, which deters informed traders from leveraging their information on the market. Informed trading plays an important role in improving market efficiency (Shleifer and Vishny, 1997). Some naïve investors may cause prices to deviate from their fundamental values, but informed traders and arbitrageurs can choose to take positions against noise traders and bring prices back to their fundamental values. Theoretically, arbitrage requires no capital and entails no risk (Sharpe and Alexander, 1990). Shleifer and Vishny (1997), however, argue that arbitrage is often risky in practice and that professional arbitrageurs can avoid extremely volatile positions. That is, professional arbitrageurs may quit the market when it is highly volatile, which can cause asset prices to deviate from fundamental values for an appreciable length of time (Gemmill and Thomas, 2002).

De Long et al. (1990) point out that the unpredictability of noise traders' future beliefs creates a risk in the price of the asset that deters rational arbitrageurs from aggressively taking positions against them. Barberis et al. (1998) argue that arbitrage is limited because movements in investor sentiment are in part unpredictable, and therefore arbitrageurs betting against mispricing bear the risk that investor sentiment may become more extreme causing prices to move even further away from fundamental values. Since high investor sentiment increases noise trading and noise trader risk, the theory of limits to arbitrage suggests that informed trading tends to be less active during high sentiment periods.

Furthermore, noise trading and noise trader risk can be closely related to trading costs. Both theoretical and empirical studies find that uncertainty measured by price volatility is positively related to the bid–ask spread (Roll, 1984; Cohen et al., 1986; French and Roll, 1986; Amihud and Mendelson, 1987; Glosten, 1987), which is one of the most important measures of trading costs in the literature. Amihud and Mendelson's (1987) theoretical model suggests a positive relation between the measured return volatility and bid–ask spread. Roll (1984), French and Roll (1986), and Glosten (1987) empirically show a positive relation between the standard deviation of transaction price changes and the bid–ask spread. McInish and Wood (1992) analyze the intraday pattern in bid–ask spreads for NYSE stocks and find that bid–ask spreads are positively related to risk. Consistent with the implications of a simple asymmetric information model for the bid–ask spread, Bollerslev and Melvin (1994) provide evidence showing that the bid–ask spread is positively related to the underlying uncertainty.

[2] Chan (1992) and Frino et al. (2000) analyze the lead–lag relation between the cash market and stock index futures market and show that the leading position of the futures market is in fact due to the information advantage of the futures market, not due to the infrequent trading problem of component stocks.

Given that increased noise trading and noise trader risk result in higher price volatility and larger uncertainty during high sentiment periods, we expect a positive relation between investor sentiment and the bid–ask spread. As a result, the increased bid–ask spread makes it more costly for informed traders to exploit their information during high sentiment periods.

Overall, as high sentiment increases noise trading and hence increases price volatility and the bid–ask spread on the futures market, based on the theory of limits to arbitrage and the trading cost hypothesis, we posit that informed traders become less willing to trade on the futures market with the increased trading risk and costs during high sentiment periods. Consequently, compared to low sentiment periods, the futures market reacts to new information relatively slowly during high sentiment periods. The asymmetric lead–lag relation between the futures and spot markets changes with investor sentiment over time. We thus state the following hypothesis:

Hypothesis 1. *The short-run leading role of the futures market is weakened during high investor sentiment periods.*

Next, if the variation in the temporal lead–lag relation between the futures and spot markets is largely due to the variation in informed trading on the futures market and if high sentiment decreases informed trading on the futures market persistently, we expect to observe not only changes in the temporal lead–lag relation between the two markets but also changes in the long-run contribution to price discovery process across markets. The lead–lag relation, which usually manifests within minutes, only illustrates which market reflects new information faster in a relatively short horizon. Informed trading, however, also affects prices in the long-run equilibrium (Hasbrouck, 1995; Gonzalo and Granger, 1995). As informed traders are less willing to leverage their information advantages on the futures market during high sentiment periods, we expect that futures prices become relatively less informative and contribute less to the price discovery process in equilibrium. Thus, we state our second hypothesis:

Hypothesis 2. *The prices on the futures market contribute less to the price discovery process in the long-run equilibrium during high investor sentiment periods.*

3. Data and measurement of variables

3.1. Data

The data used in this study consist of intraday trade and quote prices of three ETFs and their corresponding E-mini index futures, which include the S&P 500 ETFs and S&P 500 E-mini futures, the Nasdaq 100 ETFs and Nasdaq 100 E-mini futures, and the DJIA ETFs and DJIA E-mini futures. These three ETFs and futures price pairs are examined because they are among the most active index-tracking ETFs and futures contracts on the market. By studying these price series, we can better control for the potential biases caused by the infrequent trading problem.

Our sample period is from January 1, 2002 to December 31, 2010.[3] The sample period ends in December 2010 due to the availability of sentiment index from Baker and Wurgler (2006). Specifically, Baker and Wurgler construct the sentiment index by taking the first principal component of six sentiment-related measures from January 1965 to December 2010. Although they have recently updated the index to 2014, the index is different from the original one because they withdraw one of the six measures, NYSE turnover, from the calculation. To test the robustness of our results, we perform additional tests using another sentiment index from Huang et al. (2015), in which they use partial least square (PLS) method to calculate the PLS sentiment index with the same six sentiment-related measures as in Baker and Wurgler. For consistency and comparability between our main and robustness tests, we use the original Baker and Wurgler sentiment index for relevant tests with a sample period ending in December 2010.[4]

The tick-by-tick quote data of ETFs are obtained from the Trade and Quote database. We take the midpoint of the quoted bid and ask prices as the proxy for the fundamental values of the ETFs and, following Hasbrouck (2003), use regular quotes on the primarily listed market of the ETFs.[5] Regular trading hour is between 9:30 AM and 4:00 PM EST.[6]

Because the quote data for index futures are unavailable, we use trading prices of the futures contracts. The corresponding index futures contracts, including the E-mini versions of the S&P 500, Nasdaq 100, and DJIA index futures, are obtained from the Chicago Mercantile Exchange (CME). These E-mini futures contracts are quite active. As Hasbrouck (2003) shows, they dominate the price discovery process on the S&P 500 and Nasdaq 100 index markets. We use the nearby futures contracts because they are the most actively traded contracts. To construct a continuous time series for the futures prices, we replace the prices of the nearby contract by those of the first deferred contract, once the daily trading volume of the first deferred contract exceeds that of the nearby contract.

The prices of ETFs and futures contracts are not uniformly spaced in time. To assess the degree of comovement among the prices between different markets, we follow the method in Chan (1992) to synchronize price pairs. We construct a minute-by-minute data set for each price series. The daily trading hours are from 9:30 a.m. to 4:00 p.m., which contain 390 minute-by-minute intervals on each trading day. In each one-minute interval, we identify the last price observation. If no price is observed within that one-minute span, we use the price of the previous minute instead.[7] Returns are calculated as differences in log prices.[8] There are 879,916; 880,802; and 847,103 one-minute price pairs for the S&P 500, Nasdaq 100, and DJIA ETFs and futures, respectively, in our nine-year sample period.[9] We obtain trading volume for the ETFs and futures from the Trade and Quote database and the CME, respectively.[10]

3.2. Measurement of variables

We measure investor sentiment by adopting the monthly market-based sentiment index constructed by Baker and Wurgler (2006). Baker and Wurgler form their sentiment index by tak-

[3] For the DJIA ETFs and DJIA E-mini futures price pairs, the sample period is from May 1, 2002 to December 31, 2010 because the trading of the DJIA E-mini futures starts on May 1, 2002.

[4] The additional tests show consistent results and are available in the Internet Appendix.

[5] For the S&P 500 ETFs and DJIA ETFs, we use the AMEX quotes before September 30, 2008 and the NYSE quotes after October 1, 2008 since the AMEX is merged by the NYSE after October 1, 2008. Similarly, for the Nasdaq 100 ETFs, the quotes from the AMEX before November 30, 2004 and the quotes from NASDAQ after December 1, 2004 are used because the Nasdaq 100 ETFs transferred its listing from the AMEX to NASDAQ on December 1, 2004.

[6] To avoid contaminating effects due to the financial crisis surrounding the end of 2008, we repeat our tests with data before September 1, 2008. The empirical results are qualitatively similar and available upon request.

[7] Because our sample ETFs and E-mini futures are actively traded, we use the price of the previous minute in less than 1% of the sample.

[8] To avoid data errors, the minute-by-minute returns are winsorized at the 0.5 and 99.5 percentiles. Empirical results are qualitatively similar with or without winsorization.

[9] Due to occasional market suspensions and trading halts, we remove sample days in which less than 120 one-minute quote or trade prices can be identified.

[10] The trading volume data for futures are available since July 1, 2003.

ing the first principal component of six sentiment-related measures. The six measures are the closed-end fund discount, the number and the first-day returns of IPOs, NYSE turnover, the equity share in total new issues, and the dividend premium. The principal component analysis filters out idiosyncratic noise in the six measures and captures their common component. Similar to Yu and Yuan (2011), we define high (low) sentiment periods as those months in which Baker and Wurgler's investor sentiment index is above (below) its median during our sample period from January 2002 to December 2010. We also create high (low) sentiment dummies in which the investor sentiment index is above (below) its 75th or 50th (25th) percentile in the empirical tests.[11]

Huang et al. (2015) suggest that the first principal component of Baker and Wurgler's (2006) six sentiment-related proxies is problematic because, given true but unobservable investor sentiment, the proxies may include common approximation errors are not relevant to investor sentiment. Thus, the first principal component can potentially contain a substantial amount of common approximation errors that are not relevant to investor sentiment. To address this concern, Huang et al. use the PLS method to align the investor sentiment measure and propose a more efficient and powerful (in terms of predicting stock returns) investor sentiment index. Following Huang et al., we use the PLS sentiment index to test the robustness of our study, and the Internet Appendix reports the results.[12]

To investigate the impact of sentiment on price volatility, we calculate the daily realized volatility for ETFs and futures as

$$\text{Realized Volatility } (RV_t) = \sqrt{\sum_{i=1}^{m} (r_i)^2} \times 10,000, \tag{1}$$

where m is the number of one-minute intervals during the regular trading hours on day t, and r_i is the ith one-minute return on a trading day. Furthermore, we use quoted spreads to measure the trading costs of the ETFs, which are defined as the differences between quoted ask and bid prices. Because the quote data are not available for the E-mini futures, we use the price reversal method, suggested by Wang et al. (1994), to calculate the daily estimated bid–ask spread for the futures.[13]

4. Empirical methods

4.1. Reduced form models for volatility and bid–ask spreads

We use two reduced form equations to estimate the potential impacts of investor sentiment on price volatility and bid–ask spreads with other control variables.[14] The two empirical models are specified as follows, beginning with the price volatility equation;

$$RV_t = \alpha_0 + \beta_0 D_t^{high-sent} + \sum_{i=1}^{n} \gamma_{0i} RV_{t-i} + \sum_{i=1}^{n} \theta_{0i} TV_{t-i} + \sum_{i=1}^{n} \delta_{0i} BAS_{\ t-i} + \varepsilon_t, \tag{2}$$

where RV_t is the realized volatility on day t, and $D_t^{high-sent}$ is a dummy variable of high sentiment on day t, which is equal to 1 if the sentiment index is above its 75th percentile during our sample period, and zero otherwise. RV_{t-i} is the realized volatility on day $t - i$ to take account of the cluster effects of volatility, TV_{t-i} is the daily trading volume on day $t - i$, and $BAS_{\ t-i}$ is the percentage quoted bid–ask spread for the ETFs and the percentage estimated bid–ask spread for the futures on day $t - i$. ε_t is the error term. The percentage bid–ask spread is defined as the spread scaled by price. We expect price volatility to be positively related to investor sentiment. A significantly positive β_0 would support the assertion that high sentiment introduces more noise traders to the market and makes the market more volatile.

Similarly, the bid–ask spread equation is specified as

$$BAS_t = \alpha_1 + \beta_1 D_t^{high-sent} + \sum_{i=1}^{n} \gamma_{1i} RV_{t-i} + \sum_{i=1}^{n} \theta_{1i} TV_{t-i} + \sum_{i=1}^{n} \delta_{1i} BAS_{\ t-i} + \varepsilon_t. \tag{3}$$

A significantly positive β_1 would support the assertion that high sentiment increases trading costs proxied by the bid–ask spreads. The ordinary least square method is used to estimate the parameters with Newey–West heteroscedasticity and autocorrelation consistent covariance matrix used to obtain consistent estimates of the standard errors of the parameters.

4.2. Time-varying lead–lag relation between the futures and spot markets

Next, we use the vector error correction model (VECM) to investigate the time-varying lead–lag relation between the spot and futures markets. Index tracking ETFs and their corresponding futures are based on the same underlying assets, so we assume that they share the same implicit efficient price component. Therefore, prices on these two markets form a cointegration system.[15] If two price series are cointegrated, the Granger representation theorem (Engle and Granger, 1987) suggests that we should use VECM to model the temporal relation between the price changes of these two series. Our VECM for the spot and futures price changes is specified as

$$\Delta P_t = \mu + \sum_{i=1}^{k} (A_{1i} + A_{2i} \times sent\_dum) \Delta P_{t-i} + \gamma z_{t-1} + \varepsilon_t, \tag{4}$$

where Δ is the difference operator, P_t is a (2×1) vector of log prices on the two markets, μ is a (2×1) vector of constants, $A_{1i}\, and\, A_{2i}$ are both (2×2) matrices of autoregressive coefficients, $sent\_dum$ is a dummy variable of investor sentiment, k is the number of lags, γ is a (2×1) vector of coefficients on the error correction terms, $z_{t-1} = \alpha' P_{t-1}$ is a scalar of error correction terms, α' is a (1×2) cointegrating vector, and ε_t is a (2×1) vector of price innovations. The coefficient, γ, on the error correction term, also called speed of adjustment, measures the price reactions to the deviations from the long-term equilibrium relation. In our case, $\Delta P_t = (\Delta \log(F_t),\ \Delta \log(S_t))'$, where F_t and S_t denote the prices for the index futures and their corresponding ETFs, respectively. We use three sets of sentiment dummies to investigate the impact of investor sentiment on the markets' lead–lag relation: (i) a high sentiment dummy that equals 1 when the sentiment index is above the 75th percentile of its distribution, and zero otherwise;

[11] The data come from Wurgler's website: http://people.stern.nyu.edu/jwurgler/.

[12] The PLS sentiment index comes from Zhou's website: http://apps.olin.wustl.edu/faculty/zhou/.

[13] The estimated bid–ask spread for the E-mini futures is calculated as follows: (a) an empirical joint price distribution of ΔP_t and ΔP_{t-1} during a daily interval is created; (b) the subset of price changes that exhibited price continuity (i.e., a positive change followed by another positive change) is discarded; (c) absolute values of the price changes that are price reversals are taken; and (d) the mean of absolute values obtained in the third step is the estimated spreads.

[14] To avoid the simultaneous equation bias, price volatility, trading volume and the bid–ask spreads are used as lagged variables in Eqs. (2) and (3). Refer to Wang and Yau (2000) for further discussions on this issue.

[15] We apply the Johansen likelihood ratio test and confirm that each matched daily price pairs of spot and future prices form a cointegrated system.

(ii) a high sentiment dummy that equals 1 when the sentiment index is above the 50th percentile of its distribution, and zero otherwise; and (iii) a low sentiment dummy, that equals 1 when the sentiment index is below the 25th percentile of its distribution, and zero otherwise.

The coefficient matrix $(A_{1i} + A_{2i} \times sent\_dum)$ in Eq. (4) reveals the temporary lead–lag relation between the two markets, in which A_{2i} is used to test Hypothesis 1 and is expected to change with different sentiment dummies. More specifically, if informed traders are reluctant to trade on the futures market during high sentiment periods due to increased risk and costs, the impact of the lagged futures returns on the ETFs returns will become smaller and less significant when sentiment is high. Such a finding will support our Hypothesis 1 that the short-run leading role of futures is weakened during high investor sentiment periods.

4.3. Information shares and factor weights

Furthermore, we adopt two price information measures, information shares (Hasbrouck, 1995) and factor weights (Gonzalo and Granger, 1995), to investigate the effect of investor sentiment on the price discovery process between the spot and futures markets in equilibrium. Hasbrouck (1995) suggests that the contribution to price discovery by each market (sharing a stochastic common trend) is defined as the variation in efficient price innovations attributable to that market's innovation. According to Hasbrouck, the efficient price v_t follows a random walk: $v_t = v_{t-1} + u_t$. The observed prices of several cointegrated markets contain the same random walk component and components incorporating the effects of market frictions.

Hasbrouck (1995) shows that the following vector moving average model (VMA) can be derived from the VECM:

$$\Delta P_t = \Psi(L)\varepsilon_t, \tag{5}$$

where $\Psi(L)$ is a polynomial in the lag operator. The VMA coefficients can be used to calculate the variance of the underlying efficient price:

$$\sigma_u^2 = \Psi\Omega\Psi', \tag{6}$$

where Ψ is a row vector composed of VMA coefficients and $\Omega = var(\varepsilon_t)$.

Using the Cholesky factorization to transform Ω into a lower triangular matrix F, and $\Omega = FF'$, the information share of market j is calculated as:

$$IS_j = \frac{(\Psi F)_j^2}{\sigma_u^2}, \tag{7}$$

where $(\Psi F)_j$ is the jth element of the row matrix ΨF. The larger information share of the jth market, the more predominant force it has in setting the common efficient price. By permuting the order of the market prices, Eq. (7) provides an upper and a lower bound for the information shares of each market. We compute the information shares for our three ETFs–futures pairs each day and use the midpoint of the upper bound and lower bound as the measure for information shares.

In general, the market with larger information shares is considered to contribute more to the discovery of the long-run equilibrium price. If the information shares of futures, IS_f, is negatively related to investor sentiment, then Hypotheses 2 is supported, which indicates that the prices on the futures market contribute less to the price discovery process in equilibrium during high investor sentiment periods. Because the literature shows that information shares are related to liquidity and volatility (Eun and Sabherwal, 2003; Ates and Wang, 2005), the regression model for IS_f is specified as

$$IS_{f,t} = \alpha_2 + \beta_2 D_t^{high-sent} + \gamma_2 liquidity_t + \theta_2 volatility_t + \varepsilon_t, \tag{8}$$

where $IS_{f,\,t}$ is the information shares of the futures contracts on day t; $D_t^{high-sent}$ is a dummy variable of high sentiment on day t, which is equal to 1 if the sentiment index is above its 75th percentile during our sample period, and zero otherwise; $liquidity_t$ is the liquidity measure on day t; and $volatility_t$ is the realized volatility on day t. We use three different proxies for liquidity: market share (MS), spread ratio (SR), and trading volume (TV). MS is the market share of futures, defined as the ratio of dollar volume of futures to the sum of dollar volume of futures and the corresponding ETFs. It is calculated as

$$MS_t = \frac{Dollar\ volume\ of\ futures\ on\ day\ t}{Dollar\ volume\ of\ futures\ and\ ETFs\ on\ day\ t}. \tag{9}$$

SR is the ratio of the bid–ask spread of the ETFs to the bid–ask spread of the futures, and TV is the daily trading volume of the futures contracts. The literature shows that liquidity is positively related to the contribution of price discovery (Eun and Sabherwal, 2003; Ates and Wang, 2005). More important, we expect β_2 in Eq. (8) to be significantly negative, which would support Hypothesis 2.

In a cointegrated system such as that in Eq. (4), Gonzalo and Granger (1995; hereafter GG) also propose a methodology to decompose the vector of market prices into permanent and transitory components:

$$P_t = f_t i_2 + z_t, \tag{10}$$

where P_t is a (2×1) vector of log prices on the futures and spot markets on day t, f_t is a scalar of common long-memory component, z_t is a (2×1) transitory component, and i_2 is a (2×1) unit vector. The Johansen's maximum likelihood framework suggests that the common long-memory factor can be estimated as $f_t = \gamma_\perp' P_t$, where $\gamma_\perp'$ is a (1×2) vector, which is orthogonal to the vector of speed of adjustment coefficients, γ, on the error correction term in Eq. (4). The common factor has been interpreted as the implicit efficient price, which is common to the related market prices.

The normalized GG factor weights, $\gamma_\perp'$, are used as an alternative measure of the contribution to price discovery by each related market. The GG factor weights are summed to 1 and a larger factor weight of the jth market price suggests that the jth market price makes a larger contribution to the price discovery process in equilibrium. We expect that investor sentiment has a negative impact on the GG factor weights of the futures market.

5. Empirical results

5.1. Investor sentiment, price volatility, bid–ask spreads, and the lead–lag relation

This section presents empirical results on the impact of investor sentiment on price volatility, the bid–ask spread, and the short-run lead–lag relation between the futures and spot markets. Table 1 reports summary statistics for the investor sentiment index, the minute-by-minute returns, daily average bid–ask spread, percentage bid–ask spread, trading volume, and realized volatility of ETFs and futures.[16] Our nine-year sample period from January 2002 to December 2010 includes 108 monthly observations for the sentiment index from Baker and Wurgler (2006). The sentiment index

[16] We use the orthogonal investor sentiment index from Baker and Wurgler (2006) who regress each of the six raw proxies for sentiment on industrial production index, growth in consumer durables, nondurables, and services, and a dummy variable for NBER recession. They argue that the residuals from these regressions are cleaner proxies for investor sentiment. The orthogonal investor sentiment index is the first principal component of the correlation matrix of these six residuals.

Table 1
Summary statistics.
This table presents summary statistics of the monthly investor sentiment index, the minute-by-minute return series, daily bid–ask spread, percentage bid–ask spread, trading volume, and realized volatility for the S&P 500, Nasdaq 100, and DIJA ETFs and their corresponding E-mini futures. Sentiment index is the orthogonal investor index, which is orthogonal to macroeconomic factors, from Baker and Wurgler (2006). The ETFs prices are the mid-points of quoted bid and ask prices in their primarily listing market. The futures prices are the transaction prices from the CME. The minute-by-minute return series are calculated as differences in log prices. The bid–ask spread of ETFs is the daily average of quoted spread, and the bid–ask spread of futures is the spread estimated by the price reversals method with intraday transaction data. The percentage bid–ask spread is defined as the spread scaled by price. Trading volume is the sum of daily trading. Realized volatility is calculated by one-minute returns within a trading day as defined in Eq. (1). The ETFs data are from the Trade and Quote database, and the E-mini futures data are from the Chicago Mercantile Exchange (CME). The sample period is 1/1/2002 to 12/31/2010 for the sentiment index and the return pairs for the S&P 500 and Nasdaq 100 indexes. The sample period is 5/1/2002 to 12/31/2012 for the DIJA index because the DIJA E-mini futures started being listed on the CME from 5/1/2002.

Variables	Mean	Median	SD	Skewness	Kurtosis	# Obs.
Sentiment index	−0.077	−0.060	0.319	−0.013	0.917	108
S&P 500 ETFs						
Returns (%)	0.000	0.000	0.051	0.017	3.828	877,652
Bid–ask spread	0.044	0.042	0.025	0.994	3.188	2112
Percentage bid–ask spread (%)	0.039	0.033	0.025	1.315	2.366	2112
Trading volume (million)	19.647	6.397	29.761	2.517	7.282	2267
Realized volatility (%)	0.863	0.762	0.380	1.479	2.627	2110
S&P 500 Futures						
Returns (%)	0.000	0.000	0.053	0.002	3.349	877,652
Bid–ask spread	0.251	0.250	0.003	2.286	15.080	2150
Percentage bid–ask spread (%)	0.022	0.021	0.003	0.471	0.131	2150
Trading volume (million)	1.179	0.979	0.717	1.094	1.611	1931
Realized volatility (%)	0.903	0.796	0.394	1.411	2.243	2110
Nasdaq 100 ETFs						
Returns (%)	0.000	0.000	0.066	0.010	3.027	878,537
Bid–ask spread	0.015	0.011	0.008	2.459	7.930	2266
Percentage bid–ask spread (%)	0.045	0.027	0.035	1.996	3.274	2266
Trading volume (million)	39.562	30.653	33.911	1.996	6.473	2267
Realized volatility (%)	1.187	1.031	0.544	1.427	2.456	2265
Nasdaq 100 Futures						
Returns (%)	0.000	0.000	0.063	−0.003	2.870	878,537
Bid–ask spread	0.381	0.298	0.119	0.094	−1.949	2314
Percentage bid–ask spread (%)	0.023	0.024	0.009	0.413	−0.983	2314
Trading volume (million)	0.279	0.270	0.106	0.669	2.936	1931
Realized volatility (%)	1.134	0.995	0.503	1.598	3.431	2265
DIJA ETFs						
Returns (%)	0.000	0.000	0.051	0.012	4.135	844,922
Bid–ask spread	0.070	0.076	0.041	0.631	1.696	2094
Percentage bid–ask spread (%)	0.068	0.072	0.040	0.568	0.728	2094
Trading volume (million)	1.732	0.967	2.244	3.063	13.271	2267
Realized volatility (%)	0.859	0.760	0.386	1.399	2.330	2010
DIJA Futures						
Returns (%)	0.000	0.000	0.051	0.016	4.392	844,922
Bid–ask spread	1.242	1.116	0.377	3.434	11.924	2031
Percentage bid–ask spread (%)	0.012	0.010	0.005	2.581	7.121	2031
Trading volume (million)	0.113	0.101	0.057	1.380	2.963	1909
Realized volatility (%)	0.843	0.721	0.425	1.517	2.542	2010

is a standardized statistics with zero mean and unit variance. In Table 1, the mean and median of the sentiment index are both close to zero, which indicates that investor sentiment during our sample period is not leaning toward either high or low levels.

Table 1 shows that each of our sample price pairs behaves similarly. This result provides evidence that our sample ETFs and futures based on the same underling assets are closely and highly related. For instance, the mean and median returns of S&P 500 ETFs and futures are all close to zero with similar skewness and kurtosis; the realized volatility of S&P 500 ETFs and futures are 0.863 and 0.903, respectively.[17] Moreover, the trading volume of our sample ETFs and futures contracts is quite high, suggesting that they are actively traded. For example, the daily average number of ETF shares and futures contracts traded on the S&P 500 is 19.65 million and 1.18 million, respectively. As a result, we can better control for the potential biases caused by the infrequent trading problem in our empirical tests.

Table 2 presents the regression results of the relation between investor sentiment and price volatility. We use the minute-by-minute realized volatility as a proxy for price volatility and then regress it on the high sentiment dummy and other control variables. The high sentiment dummy equals 1 when the sentiment index is above its 75th percentile during our sample period and zero otherwise. After controlling for lagged realized volatility, lagged trading volume, and lagged bid–ask spread, we find that the sentiment dummy has a significantly positive impact on volatility for all sample ETFs and futures returns. The results are robust with control variables of various lags.

For example, with lag 1 control variables, the coefficients on the high sentiment dummy in Models 1 and 2 of Table 2 are, respectively, 4.81 and 7.50, significant at the 1% level, with S&P 500 ETFs and futures volatility as the dependent variables, respectively. This shows that high investor sentiment significantly increases price volatility. The coefficients on the sentiment dummy remain significantly positive in Models 3–6 when lag 2 and 3 control variables are added. Similar results are found for the Nasdaq 100 and DJIA

[17] We are not concerned about ex-dividend day drops for the ETFs returns because for each day we use the minute-by-minute price series within trading hours to calculate the returns and realized volatility, and ex-dividend adjustments usually take place before trading hours.

Table 2
Realized volatility and investor sentiment.
This table presents the regression results of realized volatility for the S&P 500, Nasdaq 100, and DJIA ETFs and E-mini futures. We regress realized volatility on the high sentiment dummy, lagged realized volatility (*RV*), lagged trading volume (*TV*), and lagged percentage bid–ask spread (*BAS*). The prefix LAG*n* means *n*-day lagged terms. The high sentiment dummy is equal to 1 if the sentiment index is above its 75th percentile in our sample period, and zero otherwise. Realized volatility is calculated by one-minute returns within a trading day as

$$\textit{Realized Volatility } (RV_t) = \sqrt{\sum_{i=1}^{m} (r_i)^2} \times 10,000,$$

where m is the number of one-minute intervals during the regular trading hours on day t, and r_i is the ith one-minute return on a trading day. Trading volume is the sum of daily trading. The percentage bid–ask spread of ETFs is the daily average of quoted spread divided by the ETFs prices. The percentage bid–ask spread of the futures is the spread estimated by the price reversals method with intraday transaction data, and it is also divided by the futures prices. The sample period is 7/1/2003 to 12/31/2010. p-values are in parentheses. In all tests, ***, **, and * signify statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable: realized volatility	S&P500						Nasdaq100						DJIA					
	(1) ETFs	(2) Futures	(3) ETFs	(4) Futures	(5) ETFs	(6) Futures	(7) ETFs	(8) Futures	(9) ETFs	(10) Futures	(11) ETFs	(12) Futures	(13) ETFs	(14) Futures	(15) ETFs	(16) Futures	(17) ETFs	(18) Futures
Intercept	9.553***	−16.983***	7.416***	−11.111***	6.328***	−7.520**	11.903***	17.432***	8.464***	13.849***	7.632***	13.323***	11.478***	−10.838***	9.323***	−6.273	8.104***	0.180
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.15)	(0.00)	(0.97)
High Sent. Dummy	4.814***	7.502***	3.919***	5.018***	3.610***	4.511***	3.891***	5.652***	3.418***	4.454***	3.635***	4.641***	4.756***	7.214***	3.959***	5.192***	3.401***	4.731***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LAG1_RV	0.853***	0.862***	0.632***	0.550***	0.593***	0.483***	0.907***	0.961***	0.605***	0.602***	0.549***	0.535***	0.835***	0.873***	0.613***	0.585***	0.584***	0.540***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LAG2_RV			0.256***	0.342***	0.192***	0.244***			0.341***	0.373***	0.259***	0.308***			0.259***	0.327***	0.189***	0.241***
			(0.00)	(0.00)	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)
LAG3_RV					0.126***	0.188***					0.150***	0.138***					0.128***	0.187***
					(0.00)	(0.00)					(0.00)	(0.00)					(0.00)	(0.00)
LAG1_TV ($\times 10^{-7}$)	0.119***	1.576	0.091**	57.618***	0.115***	71.459***	−0.042	−530.008***	0.078**	−34.774	0.114***	62.090	1.622***	−369.901**	1.209***	367.134*	1.313***	486.166**
	(0.00)	(0.88)	(0.03)	(0.00)	(0.01)	(0.00)	(0.11)	(0.00)	(0.05)	(0.67)	(0.01)	(0.47)	(0.00)	(0.03)	(0.00)	(0.08)	(0.00)	(0.02)
LAG2_TV ($\times 10^{-7}$)			0.002	−48.187***	0.098**	−1.939			−0.128***	−401.612***	−0.045	−225.362***			0.089	−649.626***	0.779*	−242.040
			(0.97)	(0.00)	(0.05)	(0.92)			(0.00)	(0.00)	(0.32)	(0.01)			(0.83)	(0.00)	(0.10)	(0.28)
LAG3_TV($\times 10^{-7}$)					−0.151***	−67.030***					−0.122***	−271.767***					−1.310***	−732.610***
					(0.00)	(0.00)					(0.00)	(0.00)					(0.00)	(0.00)
LAG1_BAS ($\times 10^{4}$)	−0.175	12.682***	−0.160	31.860***	−0.144	38.961***	−0.324	0.252	−0.694	25.929***	−0.771	32.084***	−0.218***	21.503***	−0.122	18.669***	−0.116	19.273***
	(0.40)	(0.00)	(0.45)	(0.00)	(0.50)	(0.00)	(0.57)	(0.73)	(0.33)	(0.00)	(0.31)	(0.00)	(0.01)	(0.00)	(0.14)	(0.01)	(0.16)	(0.01)
LAG2_BAS ($\times 10^{4}$)			−0.109	−23.372**	−0.180	−17.384			0.230	−25.888***	0.187	−22.109***			−0.138*	−4.796	−0.100	−5.459
			(0.60)	(0.03)	(0.40)	(0.21)			(0.75)	(0.00)	(0.81)	(0.01)			(0.09)	(0.49)	(0.23)	(0.49)
LAG3_BAS ($\times 10^{4}$)					−0.001	−15.317					−0.039	−10.012					−0.079	−7.694
					(0.99)	(0.17)					(0.96)	(0.16)					(0.34)	(0.30)
Adj. R-Square	0.84	0.86	0.86	0.87	0.86	0.88	0.79	0.82	0.81	0.84	0.81	0.84	0.83	0.87	0.84	0.88	0.84	0.88
Obs.	1846	1846	1803	1803	1760	1760	1848	1848	1806	1806	1764	1764	1864	1864	1841	1841	1818	1818

Table 3
Bid–ask spread and investor sentiment.
This table presents the regression results of bid–ask spread for the S&P 500, Nasdaq 100, and DJIA ETFs and E-mini futures. We regress bid–ask spread on the high sentiment dummy, lagged realized volatility (*RV*), lagged trading volume (*TV*), and lagged bid–ask spread (*BAS*). The prefix LAG*n* indicates the *n*-day lagged terms. The high sentiment dummy is equal to 1 if sentiment index is above its 75th percentile in our sample period, and zero otherwise. Realized volatility is the sum of one-minute squared returns within a trading day. Trading volume is the sum of daily trading. The bid–ask spread of ETFs is the daily average of quoted spread. The bid–ask spread of futures is the spread estimated by the price reversals method with intraday transaction data. The percentage bid–ask spread is defined as the spread scaled by price. The sample period is 7/1/2003 to 12/31/2010. *p*-values are in parentheses. In all tests, ***, **, and * signify statistical significance at the 1%, 5%, and 10% levels, respectively.

	Bid–ask spread ($\times 10^4$)						Percentage bid–ask spread ($\times 10^6$)					
	S&P500		Nasdaq100		DJIA		S&P500		Nasdaq100		DJIA	
	ETFs	Futures	ETFs	Futures	ETFs	Futures	ETFs	Futures	ETFs	Futures	ETFs	Futures
Intercept	43.85***	106.23***	4.20***	42.83***	109.88***	52.87***	23.38***	49.49***	8.11***	70.95***	75.21***	36.80***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
High Sent. Dummy	8.04*	52.35***	2.46**	12.05***	40.52***	29.29***	1.88	7.70*	4.96*	14.81***	10.06	13.37***
	(0.06)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)	(0.63)	(0.10)	(0.09)	(0.01)	(0.28)	(0.01)
LAG1_RV ($\times 10^2$)	0.23***	1.03***	0.01	0.24***	0.41**	0.66***	−0.05	0.75***	0.02	0.75***	0.18	1.37***
	(0.01)	(0.00)	(0.24)	(0.00)	(0.02)	(0.00)	(0.52)	(0.01)	(0.74)	(0.00)	(0.25)	(0.00)
LAG1_TV ($\times 10^{-3}$)	−0.82***	78.04	−0.05**	−101.10	−19.59***	254.57	−0.16	−86.22*	−0.10	−1449.80***	−9.62***	−2462.88***
	(0.00)	(0.14)	(0.02)	(0.40)	(0.00)	(0.70)	(0.19)	(0.08)	(0.18)	(0.00)	(0.00)	(0.00)
LAG1_BAS	0.88***	0.56***	0.97***	0.67***	0.82***	0.66***	0.95***	0.07***	0.98***	0.08***	0.88***	0.06***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Adj. *R*-Square	0.89	0.73	0.97	0.69	0.78	0.76	0.92	0.80	0.97	0.81	0.81	0.83
Obs.	2067	1749	2216	1889	1978	1737	2067	1749	2216	1889	1978	1737

ETFs and futures volatilities in Table 2.[18] These results suggest that investor sentiment has a positive impact on price volatility and are consistent with Barberis et al. (1998), Brown (1999), Yu and Yuan (2011), and Stambaugh et al. (2012), who argue that high sentiment introduces more noise trading into the market and in turn makes the markets noisier and riskier.

Interestingly, Table 2 shows that the coefficients on the high sentiment dummy in the futures (ETFs) regressions are moderately higher (lower). For example, the coefficients on the high sentiment dummy for the S&P 500 in Models 1, 3, and 5 are 4.81, 3.91, and 3.61, respectively; the coefficients in Models 2, 4, and 6 are consistently higher at 7.50, 5.01, and 4.51, respectively. The Nasdaq 100 and DJIA estimations display similar patterns. The empirical evidence shows that investor sentiment in terms of price volatility has a greater effect on the futures market than on the spot market and that the futures market becomes even more volatile during high sentiment periods. Thus, informed traders bear relatively more risks when they utilize their information on the futures market during high sentiment periods.

We conduct two robustness checks. First, we replace the contemporary high sentiment dummy with one-month lagged high sentiment dummy. Second, instead of the sentiment index from Baker and Wurgler (2006), we use the PLS sentiment index from Huang et al. (2015) to set the high sentiment dummy. The results (please see Tables A1 and A2 in the Internet Appendix) are similar to those in Table 2, indicating that investor sentiment has a positive impact on price volatility.

Table 3 presents the results of regressing the bid–ask spread and the percentage bid–ask spread on the high sentiment dummies and other control variables and shows that investor sentiment has a significantly positive impact on both the bid–ask spread and the percentage bid–ask spread. When the dependent variables are the bid–ask spreads of the S&P 500, Nasdaq 100, and DJIA ETFs and futures, the coefficients on the high sentiment dummy are all positive and statistically significant at the 10% level. Take the bid–ask spread of the Nasdaq 100 ETFs and futures as an example: The coefficients on high sentiment dummy are 2.46 and 12.05, respectively, significant at the 5% level. From Table 3, we further find that the high sentiment dummies have a positive impact on the percentage bid–ask spread, with all six coefficients being positive and four out of them being statistically significant at the 10% level.[19] The evidence suggests that investor sentiment has a positive impact on both ETFs and futures bid–ask spread.

Table 3 also shows that the impact of investor sentiment on the futures market is somewhat more significant than that on the spot market. Specifically, the coefficients on the high sentiment dummy for the S&P 500, Nasdaq 100, and DJIA ETFs in the percentage bid–ask spread regressions are 1.88, 4.96, and 10.06, respectively; the coefficients on the high sentiment dummy for the S&P 500, Nasdaq 100, and DJIA futures are larger and more significant at 7.70, 14.81, and 13.37, respectively. These results are consistent with the pattern reported in Table 2, indicating that investor sentiment is likely to have a greater impact on the futures market. Our findings in the bid–ask spread regressions suggest that informed traders incur more costs on the futures market during high sentiment periods.[20]

Again, as robustness checks, we use the previous month sentiment dummy and the PLS sentiment index to investigate the impact of investor sentiment on the bid–ask spreads. The results (please see Tables A3 and A4 in the Internet Appendix) show that investor sentiment is positively related to the bid–ask spread in both markets, along with more consistent and significant impact on the futures market. The literature commonly concludes that the bid–ask spread is an important proxy for trading costs: The larger it is, the less informed traders are willing to leverage their information.

In sum, our results in Tables 2 and 3 as well as our robustness checks in the Internet Appendix show that investor sentiment significantly increases both trading risk and trading costs on the futures market during high sentiment periods, which is likely to make informed traders less willing to leverage their information advantage on the futures market.

[18] We perform the regression specification error test, suggested by Hill et al. (2008), and find that the empirical models in Table 2 do not suffer from the model specification problem due to omitted variables.

[19] We again perform the specification error test, suggested by Hill et al. (2008), and show that the empirical models in Table 3 do not suffer from the model specification problem due to omitted variables. The results are qualitatively similar when the lag 2 and 3 control variables are added to the models.

[20] We only compare the coefficients on the high sentiment dummy in the percentage bid–ask spread regressions because the bid–ask spread of ETFs and futures are not comparable due to different price scales.

Table 4
VECM estimation for the ETFs and E-mini futures of the S&P 500 index.
This table presents the coefficient estimates of vector error correcting model (VECM) for the S&P 500 ETFs and E-mini futures. The minute-by-minute quoted bid and ask midpoints of the S&P 500 ETFs and the trading prices of the S&P 500 E-mini futures are used in estimation. The VECMs are estimated in a VAR(6) framework. To save space, this table shows only the first three coefficients on the lagged returns. We show the baseline estimation in the first two columns. Three sentiment dummies defined by different distribution percentiles are multiplied by the lagged price changes as explanatory variables for investigating sentiment effects. The sample period is 1/1/2002 to 12/31/2010. *t*-statistics are in parentheses. In all tests, *** and * signify statistical significance at the 1% and 10% level, respectively.

	Baseline		$D=1$ if sentiment $<$ 25th pctl.		$D=1$ if sentiment $>$ 50th pctl.		$D=1$ if sentiment $>$ 75th pctl.	
	ETFs	Futures	ETFs	Futures	ETFs	Futures	ETFs	Futures
D(ETFs(−1))	−0.574***	0.138***	−0.546***	0.157***	−0.599***	0.129***	−0.590***	0.133***
	(−237.60)	(52.95)	(−177.29)	(47.27)	(−188.61)	(37.46)	(−219.53)	(45.66)
D(ETFs(−2))	−0.419***	0.104***	−0.396***	0.114***	−0.442***	0.103***	−0.435***	0.102***
	(−145.61)	(33.43)	(−108.09)	(28.73)	(−116.25)	(25.20)	(−135.69)	(29.43)
D(ETFs(−3))	−0.297***	0.072***	−0.273***	0.082***	−0.321***	0.069***	−0.313***	0.069***
	(−99.23)	(22.36)	(−71.56)	(19.87)	(−81.53)	(16.27)	(−93.93)	(19.04)
D(Futures(−1))	0.579***	−0.167***	0.542***	−0.186***	0.601***	−0.165***	0.593***	−0.168***
	(258.48)	(−69.02)	(185.85)	(−59.03)	(208.79)	(−53.08)	(241.69)	(−63.42)
D(Futures(−2))	0.434***	−0.115***	0.406***	−0.123***	0.457***	−0.119***	0.454***	−0.113***
	(155.74)	(−38.09)	(113.17)	(−31.81)	(125.56)	(−30.13)	(146.99)	(−34.00)
D(Futures(−3))	0.307***	−0.082***	0.280***	−0.089***	0.331***	−0.085***	0.325***	−0.083***
	(104.41)	(−25.87)	(74.25)	(−21.93)	(85.83)	(−20.36)	(99.34)	(−23.37)
C			0.000	0.000	0.000	0.000	0.000	0.000
			(−1.14)	(−0.83)	(−1.10)	(−0.80)	(−1.10)	(−0.81)
D(ETFs(−1)) × Dummy			−0.066***	−0.046***	0.059***	0.022***	0.083***	0.021***
			(−13.17)	(−8.39)	(12.01)	(4.17)	(13.05)	(3.02)
D(ETFs(−2)) × Dummy			−0.054***	−0.024***	0.051***	−0.002	0.086***	0.005
			(−8.93)	(−3.67)	(8.79)	(−0.24)	(11.56)	(0.59)
D(ETFs(−3)) × Dummy			−0.058***	−0.026***	0.056***	0.003	0.083***	0.009
			(−9.28)	(−3.80)	(9.31)	(0.46)	(10.73)	(1.03)
D(Futures(−1)) × Dummy			0.091***	0.045***	−0.055***	−0.005	−0.079***	0.002
			(20.02)	(9.18)	(−11.96)	(−1.05)	(−13.15)	(0.33)
D(Futures(−2)) × Dummy			0.071***	0.020***	−0.053***	0.012*	−0.099***	−0.002
			(12.44)	(3.23)	(−9.33)	(1.87)	(−13.57)	(−0.28)
D(Futures(−3)) × Dummy			0.068***	0.017***	−0.055***	0.009	−0.088***	0.005
			(11.23)	(2.60)	(−9.15)	(1.38)	(−11.54)	(0.58)
Adj. R-square	0.074	0.006	0.074	0.007	0.074	0.007	0.074	0.006
Sum sq. resids	0.205	0.240	0.205	0.240	0.205	0.240	0.205	0.240
Akaike AIC	−12.403	−12.247	−12.404	−12.247	−12.403	−12.247	−12.403	−12.247

We next examine how informed traders respond to the riskier and more costly trading environment during high sentiment periods by testing the impact of investor sentiment on the lead–lag relation between ETFs and futures. Tables 4–6 report the coefficient matrixes, A_{1i} and A_{2i} in Eq. (4), and show the time-varying lead–lag relations between the ETFs and the corresponding futures for the S&P 500, Nasdaq 100, and DJIA indexes, respectively. From the baseline VECM in Tables 4–6, we find that the spot and futures markets significantly lead each other, indicating that a two-way Granger causality relation exists between ETFs and futures returns.

More important, the futures tend to lead ETFs more significantly, because the impact of lagged futures returns on the ETFs returns are larger and more significant than those of the lagged ETFs returns on the futures returns. Take the S&P 500 index as an example: From the baseline regressions in Table 4, the coefficients on the first three minutes lagged futures returns are 0.579, 0.434, and 0.307, respectively, with ETFs returns as the dependent variable. The coefficients on the first three minutes lagged ETFs returns are smaller at 0.138, 0.104, and 0.072, respectively, with futures returns as the dependent variable. Tables 5 and 6 present, respectively, the VECM estimation for the Nasdaq 100 and DJIA indexes, and the results are qualitatively similar. These results indicate that, although a two-way causal relation exists between ETFs and futures prices, the futures market assumes a more significant role in reflecting new information, which is consistent with the prior literature.

To investigate the effect of investor sentiment on the lead–lag relation, we add several sentiment dummies representing different levels of investor sentiment to the VECM. First, we add a low sentiment dummy, which is equal to 1 when the sentiment index is below the 25th percentile of its distribution, and zero otherwise. Again, take the S&P 500 index as an example: From Table 4, the coefficients on the interactions between the low sentiment dummy and the first three lagged futures returns are respectively 0.091, 0.071, and 0.068, significant at the 1% level, with ETFs returns as the dependent variable. This result indicates that the futures tend to lead ETFs more during low sentiment periods.

On the other hand, when the sentiment dummy is set with respect to the sentiment index at greater than either the 50th or the 75th percentiles, from Table 4, the coefficients on the interactions between the high sentiment dummies and the first three lagged S&P 500 futures returns are, respectively, −0.055, −0.053, and −0.055 (−0.079, −0.099, and −0.088) for the 50th (75th) percentile sentiment dummy, all significant at the 1% level, with ETFs returns as the dependent variable. These results show that the leading informational role of the futures is significantly weakened during high sentiment periods. Tables 5 and 6 report the results for the Nasdaq 100 and DJIA indexes, respectively, and the results are similar to those of the S&P 500 index in Table 4. Consistent with Hypothesis 1, these findings suggest that high investor sentiment weakens the short-run leading informational role of the futures, likely due to a riskier and more costly trading environment during high sentiment periods.

The effect of investor sentiment on the pricing dynamics between the spot and futures markets is not only statistically significant but also economically significant. For instance, from Table 4, for the model of the 75th percentile sentiment dummy, the coefficient on the first lagged futures returns is 0.593, whereas the

Table 5
VECM Estimation for the ETFs and E-mini Futures of the Nasdaq 100 Index.
This table presents the coefficient estimates of vector error correcting model (VECM) for the Nasdaq 100 ETFs and E-mini futures. The minute-by-minute quoted bid and ask midpoints of the Nasdaq 100 ETFs and the trading prices of the Nasdaq 100 E-mini futures are used in estimation. The VECMs are estimated in a VAR(6) framework. To save space, this table shows only the first three coefficients on the lagged returns. We show the baseline estimation in the first two columns. Three sentiment dummies defined by different distribution percentiles are multiplied by the lagged price changes as explanatory variables for investigating sentiment effects. The sample period is 1/1/2002 to 12/31/2010. *t*-statistics are in parentheses. In all tests, ***, **, and * signify statistical significance at the 1%, 5%, and 10% levels, respectively.

	Baseline		$D=1$ if sentiment $<$ 25th pctl.		$D=1$ if sentiment $>$ 50th pctl.		$D=1$ if sentiment $>$ 75th pctl.	
	ETFs	Futures	ETFs	Futures	ETFs	Futures	ETFs	Futures
D(ETFs(−1))	−0.540***	0.093***	−0.519***	0.098***	−0.560***	0.089***	−0.542***	0.101***
	(−265.02)	(45.64)	(−198.15)	(37.32)	(−204.82)	(32.39)	(−225.78)	(42.02)
D(ETFs(−2))	−0.382***	0.057***	−0.368***	0.054***	−0.397***	0.060***	−0.385***	0.065***
	(−161.50)	(23.95)	(−121.96)	(17.73)	(−124.50)	(18.87)	(−137.75)	(23.42)
D(ETFs(−3))	−0.268***	0.034***	−0.255***	0.035***	−0.284***	0.036***	−0.275***	0.038***
	(−110.19)	(14.16)	(−82.03)	(11.37)	(−86.82)	(10.91)	(−95.69)	(13.35)
D(Futures(−1))	0.631***	−0.119***	0.595***	−0.119***	0.652***	−0.128***	0.636***	−0.137***
	(309.64)	(−58.26)	(227.09)	(−45.38)	(238.24)	(−46.60)	(263.52)	(−56.79)
D(Futures(−2))	0.455***	−0.068***	0.426***	−0.066***	0.480***	−0.073***	0.467***	−0.078***
	(182.07)	(−27.22)	(134.81)	(−20.77)	(141.47)	(−21.48)	(156.60)	(−26.22)
D(Futures(−3))	0.315***	−0.044***	0.293***	−0.042***	0.338***	−0.048***	0.325***	−0.053***
	(119.80)	(−16.72)	(88.50)	(−12.64)	(94.42)	(−13.55)	(103.80)	(−16.82)
C	0.000**	0.000	0.000**	0.000	0.000**	0.000	0.000**	0.000
	(−2.19)	(−1.34)	(−2.28)	(−1.32)	(−2.23)	(−1.31)	(−2.24)	(−1.31)
D(ETFs(−1)) × Dummy			−0.056***	−0.013***	0.044***	0.009**	0.004	−0.030***
			(−13.34)	(−3.00)	(10.71)	(2.14)	(0.92)	(−6.64)
D(ETFs(−2)) × Dummy			−0.036***	0.009*	0.032***	−0.009*	0.007	−0.030***
			(−7.45)	(1.86)	(6.74)	(−1.84)	(1.42)	(−5.69)
D(ETFs(−3)) × Dummy			−0.037***	−0.001	0.036***	−0.004	0.021***	−0.013**
			(−7.29)	(−0.28)	(7.33)	(−0.75)	(3.92)	(−2.48)
D(Futures(−1)) × Dummy			0.109***	−0.004	−0.045***	0.017***	−0.011**	0.061***
			(25.93)	(−0.84)	(−10.97)	(4.22)	(−2.45)	(13.35)
D(Futures(−2)) × Dummy			0.092***	−0.006	−0.052***	0.011**	−0.032***	0.033***
			(17.54)	(−1.20)	(−10.26)	(2.21)	(−5.82)	(6.01)
D(Futures(−3)) × Dummy			0.068***	−0.009*	−0.048***	0.010*	−0.032***	0.028***
			(12.27)	(−1.69)	(−9.12)	(1.89)	(−5.51)	(4.81)
Adj. R-square	0.102	0.004	0.103	0.004	0.102	0.004	0.102	0.005
Sum sq. resids	0.326	0.327	0.326	0.327	0.326	0.327	0.326	0.327
Akaike AIC	−11.939	−11.937	−11.941	−11.937	−11.939	−11.937	−11.939	−11.937

coefficient on the interaction term between the first lagged futures returns and the sentiment dummy is −0.079 with ETFs returns as the dependent variable. This number indicates that, when investor sentiment is high, the coefficient on the first lagged futures returns drops by 13%. Similarly, from Table 6, for the model of the 75th percentile sentiment dummy of the DJIA index, the coefficient on the first lagged futures return drops by 23%, which is also economically significant. The coefficients on the interaction terms between the second to third lagged futures returns and the high sentiment dummy show similar patterns in signs and magnitudes.[21]

To test the robustness of our findings, we replace the sentiment index from Baker and Wurgler (2006) with the PLS sentiment index from Huang et al. (2015) to set the sentiment dummies. The Internet Appendix reports the estimation results for the S&P 500, Nasdaq 100, and DJIA ETFs and futures, and they are similar to our previous findings.

In sum, we show that investor sentiment has a significant impact on the lead–lag relation between the spot and futures markets. The leading role of the futures is significantly weakened when investor sentiment is high. These results imply that informed traders are less willing to leverage their information advantage on the futures market during high sentiment periods, when the noise trader risk and trading costs are plausibly high.

5.2. Investor sentiment and the price discovery process

In the previous section, the VECM estimation reveals that the temporary lead–lag relation between the spot and futures markets is affected by investor sentiment. Next we present the impact of investor sentiment on the information shares and GG factor weights to show whether investor sentiment affects the spot and futures prices in the long-run equilibrium. We use the intraday data to calculate the daily information shares for the ETFs and their corresponding futures. As pointed out in the research methodology section, the ordering of time series in the Hasbrouck (1995) model affects the calculations of information shares, so we focus on the average of the upper and lower bounds (i.e., the midpoint) of information shares.[22]

Table 7 reports the changes in the information shares under different sentiment regimes. The midpoints of the futures information shares are higher than those of the ETFs regardless of high or low sentiment. This pattern is consistent with our VECM results in the previous section and shows that the futures prices are, unconditionally, more informative than ETFs prices. Table 7, however, shows that as investor sentiment increases, the average information shares of the futures market decrease. For example, the information share midpoint of the S&P 500 futures during high sen-

[21] The VECMs in our tables are estimated in an AR(6) framework. To save space, we only report the coefficients on the first three lags.

[22] Baillie et al. (2002) demonstrate that the midpoint of the upper and lower bounds of information shares is a proper measure of a market's contribution to the price discovery process. Hasbrouck (2003), Ates and Wang (2005), and others have used the midpoint as a measure of a market's contribution to the price discovery process. Using a Monte Carlo simulation, Lien and Wang (2016) show that their results support the common practice of the midpoint approach in the literature.

Table 6
VECM Estimation for the ETFs and E-mini Futures of the DJIA Index.
This table presents the coefficient estimates of vector error correcting model (VECM) for the DJIA ETFs and E-mini futures. The minute-by-minute quoted bid and ask midpoints of the DJIA ETFs and the trading prices of the DJIA E-mini futures are used in estimation. The VECMs are estimated in a VAR(6) framework. To save space, this table shows only the first three coefficients on the lagged returns. We show the baseline estimation in the first two columns. Three sentiment dummies defined by different distribution percentiles are multiplied by the lagged price changes as explanatory variables for investigating sentiment effects. The sample period is 5/1/2002 to 10/31/2010. *t*-statistics are in parentheses. In all tests, *** and ** signify statistical significance at the 1% and 5% level, respectively.

	Baseline		$D = 1$ if sentiment $<$ 25th pctl.		$D = 1$ if sentiment $>$ 50th pctl.		$D = 1$ if sentiment $>$ 75th pctl.	
	ETFs	Futures	ETFs	Futures	ETFs	Futures	ETFs	Futures
D(ETFs(−1))	−0.573***	0.072***	−0.532***	0.081***	−0.616***	0.060***	−0.594***	0.063***
	(−257.60)	(30.56)	(−188.79)	(27.13)	(−207.69)	(19.11)	(−240.00)	(24.18)
D(ETFs(−2))	−0.398***	0.050***	−0.365***	0.049***	−0.437***	0.047***	−0.418***	0.048***
	(−153.54)	(18.36)	(−112.48)	(14.17)	(−125.07)	(12.79)	(−144.06)	(15.61)
D(ETFs(−3))	−0.277***	0.033***	−0.251***	0.030***	−0.308***	0.034***	−0.296***	0.032***
	(−103.35)	(11.59)	(−75.00)	(8.43)	(−84.99)	(8.88)	(−98.34)	(9.94)
D(Futures(−1))	0.572***	−0.079***	0.525***	−0.090***	0.612***	−0.074***	0.600***	−0.074***
	(271.04)	(−35.69)	(194.34)	(−31.70)	(220.56)	(−25.40)	(255.68)	(−30.02)
D(Futures(−2))	0.391***	−0.056***	0.354***	−0.054***	0.429***	−0.057***	0.417***	−0.053***
	(155.66)	(−20.99)	(112.04)	(−16.28)	(127.41)	(−16.14)	(147.50)	(−17.87)
D(Futures(−3))	0.270***	−0.037***	0.243***	−0.032***	0.300***	−0.042***	0.290***	−0.038***
	(103.00)	(−13.41)	(74.27)	(−9.30)	(84.86)	(−11.30)	(98.23)	(−12.19)
C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(−0.30)	(−0.28)	(−0.35)	(−0.28)	(−0.32)	(−0.26)	(−0.33)	(−0.27)
D(ETFs(−1)) × Dummy			−0.107***	−0.021***	0.098***	0.028***	0.117***	0.045***
			(−23.10)	(−4.23)	(21.75)	(5.86)	(20.38)	(7.53)
D(ETFs(−2)) × Dummy			−0.091***	0.006	0.086***	0.005	0.102***	0.011
			(−16.69)	(0.96)	(16.36)	(0.97)	(15.54)	(1.62)
D(ETFs(−3)) × Dummy			−0.076***	0.007	0.066***	−0.005	0.089***	0.002
			(−13.43)	(1.23)	(12.14)	(−0.91)	(13.21)	(0.23)
D(Futures(−1)) × Dummy			0.125***	0.027***	−0.093***	−0.013***	−0.138***	−0.033***
			(28.88)	(5.90)	(−21.71)	(−2.92)	(−25.36)	(−5.76)
D(Futures(−2)) × Dummy			0.106***	−0.004	−0.083***	0.005	−0.119***	−0.010
			(20.13)	(−0.77)	(−16.28)	(0.97)	(−18.82)	(−1.54)
D(Futures(−3)) × Dummy			0.079***	−0.014**	−0.062***	0.014**	−0.094***	0.006
			(14.32)	(−2.35)	(−11.82)	(2.46)	(−14.49)	(0.86)
Adj. R-square	0.085	0.002	0.086	0.002	0.086	0.002	0.086	0.002
Sum sq. resids	0.191	0.213	0.191	0.213	0.191	0.213	0.191	0.213
Akaike AIC	−12.435	−12.329	−12.436	−12.329	−12.436	−12.329	−12.436	−12.329

Table 7
Information shares and investor sentiment.
This table presents the upper bounds, lower bounds, and midpoints of information shares for the S&P 500, Nasdaq 100, and DJIA ETFs and futures. The last column shows the difference in the midpoints of information shares between high and low sentiment periods. High sentiment periods are those months when investor sentiment index is above the median. We use the minute-by-minute prices to calculate information shares each trading day during our sample period, which is 1/1/2002 to 12/31/2010 for the S&P 500 and Nasdaq 100 indexes and 5/1/2002 to 12/31/2010 for the DJIA index. In all tests, *** signifies statistical significance at the 1% level.

	Information share						
	Low sentiment			High sentiment			Diff. in mid
	Upper bound (1)	Lower bound (2)	Midpoint (3)	Upper bound (4)	Lower bound (5)	Midpoint (6)	(6) −(3)
Panel A: S&P500							
ETFs	0.792	0.122	0.457	0.843	0.114	0.478	0.021***
E-mini Futures	0.878	0.208	0.543	0.886	0.157	0.522	−0.021***
Panel B: Nasdaq100							
ETFs	0.688	0.139	0.413	0.705	0.136	0.421	0.008
E-mini Futures	0.861	0.312	0.587	0.864	0.295	0.579	−0.008
Panel C: DJIA							
ETFs	0.698	0.165	0.432	0.730	0.185	0.458	0.026***
E-mini Futures	0.835	0.302	0.568	0.815	0.270	0.542	−0.026***

timent periods is 0.021 significantly lower than that during low sentiment periods. Similar results are obtained for the Nasdaq 100 and DJIA indexes. From Table 7, we observe that the futures market becomes relatively less informative, whereas the spot market becomes relatively more informative, during high sentiment periods. [23]

We next perform multivariate regressions to investigate the relation between investor sentiment and the futures information shares with control variables, including realized volatility and liquidity measures. Table 8 shows that investor sentiment again has a significantly negative impact on the futures information shares.

[23] The results are similar when the high (low) sentiment period is defined as the sentiment index being above (below) its 75th (25th) percentile. An unreported univariate analysis shows that the correlation coefficients between investor sentiment and information shares of the S&P 500, Nasdaq 100, and DJIA futures are significantly negative.

Table 8
Regression analysis of information shares on investor sentiment.
This table presents the regression results of information shares for the S&P 500, Nasdaq 100, and DJIA E-mini futures. The dependent variable is the midpoint of information shares, which is calculated as the average of the upper bound and lower bound of information shares. We regress them on the high sentiment dummy and control variables. The high sentiment dummy is equal to 1 if sentiment index exceeds its 75th percentile during our sample period and zero otherwise. Control variables are one of the three liquidity measures and the realized volatility (*RV*). We use three different liquidity measures, which are market share (*MS*), spread ratio (*SR*), and trading volume (*TV*). *MS* is the market share of futures, defined as the ratio of dollar volume of futures to the sum of dollar volume of futures and the corresponding ETFs. *SR* is the ratio of the bid–ask spread of the ETFs to futures. *TV* is the daily trading volume of the E-mini futures contracts. The sample period is 1/1/2002 to 12/31/2010 for the S&P 500 and Nasdaq 100 indexes, and 5/1/2002 to 12/31/2010 for the DJIA index. *p*-values are in parentheses. In all tests, *** signifies statistical significance at the 1% level.

	Dependent variable: Midpoint of information shares for the E-mini futures								
	S&P500			Nasdaq100			DJIA		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.212***	0.506***	0.558***	0.457***	0.556***	0.622***	0.421***	0.563***	0.672***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
High Sent. dummy	−0.104***	−0.044***	−0.005	0.004	−0.047***	−0.048***	−0.129***	−0.114***	−0.047***
	(0.00)	(0.00)	(0.43)	(0.56)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
MS	0.366***			0.345***			0.253***		
	(0.00)			(0.00)			(0.00)		
SR		0.029***			0.048***			0.010***	
		(0.00)			(0.00)			(0.00)	
TV ($\times 10^{-7}$)			−0.953***			1.087***			−9.360***
			(0.00)			(0.01)			(0.00)
RV	0.037***	−0.016***	0.085***	−0.030***	−0.040***	−0.072***	−0.053***	−0.050***	−0.004
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.79)
Adj. *R*-Square	0.14	0.09	0.13	0.24	0.06	0.08	0.14	0.19	0.14
Obs.	1846	2207	1846	1644	1878	1644	1784	2062	1784

Table 9
Regression analysis of GG factor weights on investor sentiment.
This table presents the regression results of GG factor weights for the S&P 500, Nasdaq 100, and DJIA E-mini futures. The dependent variable is the GG factor weights, which is truncated within 0 and 1. We regress them on the high sentiment dummy and control variables. The high sentiment dummy is equal to 1 if sentiment index exceeds its 75th percentile during our sample period, and zero otherwise. Control variables are one of the three liquidity measures and the realized volatility (*RV*). We use three different liquidity measures, which are market share (*MS*), spread ratio (*SR*), and trading volume (*TV*). *MS* is the market share of futures, defined as the ratio of dollar volume of futures to the sum of dollar volume of futures and the corresponding ETFs. *SR* is the ratio of the bid–ask spread of the ETFs to futures. *TV* is the daily trading volume of the E-mini futures contracts. The sample period is 1/1/2002 to 12/31/2010 for the S&P 500 and Nasdaq 100 indexes, and 5/1/2002 to 12/31/2010 for the DJIA index. *p*-values are in parentheses. In all tests, *** and * signify statistical significance at the 1% and 10% levels, respectively.

	Dependent variable: GG factor weights for the E-mini futures								
	S&P500			Nasdaq100			DJIA		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.083	0.478***	0.538***	0.730***	0.780***	0.798***	0.083	0.416***	0.615***
	(0.38)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.37)	(0.00)	(0.00)
High Sent. Dummy	−0.123***	−0.025	−0.021	−0.025	−0.094***	−0.102***	−0.229***	−0.183***	−0.053*
	(0.00)	(0.27)	(0.44)	(0.34)	(0.00)	(0.00)	(0.00)	(0.00)	(0.07)
MS	0.494***			0.292***			0.530***		
	(0.00)			(0.00)			(0.00)		
SR		0.036***			0.064***			0.016***	
		(0.00)			(0.00)			(0.00)	
TV ($\times 10^{-7}$)			−0.188			6.070***			−18.059***
			(0.48)			(0.00)			(0.00)
RV	0.045	−0.022	−0.019	−0.153***	−0.140***	−0.249***	−0.073***	−0.059***	0.005
	(0.11)	(0.28)	(0.62)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.90)
Adj. *R*-Square	0.014	0.008	0.001	0.053	0.029	0.053	0.055	0.061	0.057
Obs.	1889	2264	1889	1890	2265	1890	1887	2181	1887

We regress the futures information shares of the S&P 500, Nasdaq 100, and DJIA indexes on the high sentiment dummies, set with respect to the sentiment index at greater than the 75th percentile, and on control variables in different model specifications and find that the coefficients on high sentiment dummies are mostly significantly negative. Take the information shares of the S&P 500 futures for example: From Table 8, the coefficients on high sentiment dummies in Models 1, 2, and 3 are −0.104, −0.044, and −0.005, respectively, for three different sets of liquidity controls, and two of them are significant at the 1% level.

Table 8 provides similar results for the Nasdaq 100 and DJIA futures, implying that futures prices contribute relatively less to price discovery in equilibrium when investor sentiment is high. The results reported in Tables 7 and 8 are in line with Shleifer and Vishny (1997), Barberis et al. (1998), and Gemmill and Thomas (2002), showing that informed traders avoid exposing themselves to extreme risk when investor sentiment is high and thus are less willing to leverage their information advantages on the futures market, which in turn makes futures prices relatively less informative during such periods.

We next use the GG factor weights as an alternative measure of information to examine the robustness of our results. The larger the GG factor weight in a market is, the more the prices in that market contribute to the price discovery process in a cointegrated system. Table 9 reports the regression results of the relation between investor sentiment and the futures GG factor weights. The coefficients on the high sentiment dummies are all negative in various model specifications, and most of them are statistically significant. For example, with the GG factor weights of the DJIA futures as the dependent variable, the coefficients on the high sentiment dummies are –0.229, –0.183, and –0.053, significant at the 10% level, in Models 7, 8, and 9, respectively. Results for the S&P 500 and Nasdaq 100 futures are similar, which again implies that the futures prices contribute less to the price discovery process when investor sentiment is high.[24]

6. Conclusion

The literature extensively shows that the futures market incorporates new information more quickly than the spot market does because lower trading costs in the futures market attract more informed traders who try to better utilize their information. Consequently, an asymmetric lead–lag relation is observed between the futures and spot markets. Some studies, however, suggest that the lead–lag relation can be time-varying if trading risk and trading costs change over time. This study empirically investigates the effect of time-varying investor sentiment on the lead–lag relation and on the price discovery process between the spot and futures markets.

Using the trade and quote data of the S&P 500, Nasdaq 100, DIJA ETFs, and their corresponding futures contracts, we first validate that investor sentiment has a positive impact on both price volatility and the bid–ask spread, which implies that informed traders bear higher trading risk and trading costs during high sentiment periods. We also find that the impact of investor sentiment on price volatility and the bid–ask spread is moderately greater and more significant in the futures market than the spot market. Based on the theory of limits to arbitrage and trading cost hypothesis, we hypothesize that informed traders become less willing to leverage their information advantages on the futures market during high sentiment periods.

Our investigation on the lead–lag relation and on the price discovery process between the futures and spot markets provides several interesting findings that are consistent with the literature and support our hypotheses. First, the short-run leading role of the futures markets becomes significantly weaker during high sentiment periods. Second, investor sentiment negatively impacts both the information shares and the GG factor weights of the futures market. Together, the results suggest that the futures prices become relatively less informative because informed traders are less willing to trade on the futures market during high sentiment periods.

Our study provides support for the theory of limits to arbitrage. That is, increased noise trader risk and trading costs during high investor sentiment periods discourage informed traders from leveraging their information advantages on the futures market. This study contributes to the literature by demonstrating empirically that the time-varying lead–lag relation between the futures and spot markets is due, at least in part, to the impact of time-varying investor sentiment. We also add to the literature by showing that investor sentiment not only affects asset prices and price volatility within a single market as prior studies suggest, but also has a significant impact on the price discovery process across informationally linked markets.

[24] To address the endogeneity concerns, we replace all control variables in Tables 8 and 9 with their lagged terms and re-estimate the coefficients. The results are both qualitatively and quantitatively similar. Moreover, the results are robust when we use the previous month sentiment dummy or the PLS sentiment index to set the dummy in the estimation. The results are available upon request.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jbankfin.2018.02.014.

References

Amihud, Y., Mendelson, H., 1987. Trading mechanisms and stock returns: an empirical investigation. J. Financ. 42, 533–553.
Ates, A., Wang, G.H.K., 2005. Information transmission in electronic versus open-outcry trading systems: an analysis of U.S. Equity index futures markets. J. Futures Mark. 25, 679–715.
Back, K., 1993. Asymmetric information and options. Rev. Financ. Stud. 6, 435–472.
Baillie, R.T., Booth, G.G., Tse, Y., Zabotina, T., 2002. Price discovery and common factor models. J. Financ. Mark. 5, 309–321.
Baker, M., Stein, J.C., 2004. Market liquidity as a sentiment indicator. J. Financ. Mark. 7, 271–299.
Baker, M., Wurgler, J., Yuan, Y., 2012. Global, local, and contagious investor sentiment. J. Financ. Econ. 104, 272–287.
Baker, M.P., Wurgler, J.A., 2006. Investor sentiment and the cross-section of stock returns. J. Financ. 61, 1645–1680.
Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. J. Financ. Econ. 49, 307–343.
Berger, D., Turtle, H.J., 2011. Emerging market crises and us equity market returns. Global Financ. J. 22, 32–41.
Black, F., 1975. Fact and fantasy in the use of options. Financ. Anal. J. 31, 36–72.
Bollerslev, T., Melvin, M., 1994. Bid–ask spreads and volatility in the foreign exchange market: an empirical analysis. J. Int. Econ. 36, 355–372.
Brown, G.W., 1999. Volatility, sentiment, and noise traders. Financ. Anal. J. 82–90.
Brown, G.W., Cliff, M.T., 2005. Investor sentiment and asset valuation. J. Bus. 78, 405–440.
Chan, K., 1992. A further analysis of the lead–lag relationship between the cash market and stock index futures market. Rev. Financ. Stud. 5, 123–152.
Cohen, K., Maier, S., Schwartz, R., Whitcomb, D., 1986. The Microstructure of Security Markets. Prentice-Hall, Englewood Cliffs, NJ.
De Long, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990. Positive feedback investment strategies and destabilizing rational speculation. J. Financ. 45, 379–395.
Easley, D., O'Hara, M., Srinivas, P.S., 1998. Option volume and stock prices: evidence on where informed traders trade. J. Financ. 53, 431–465.
Engle, R.F., Granger, C.W.F., 1987. Co-integration and error correction: representation, estimation, and testing. Econometrica 55, 251–276.
Eun, C.S., Sabherwal, S., 2003. Cross-border listings and price discovery: evidence from U.S.-listed Canadian stocks. J. Financ. 58, 549–576.
Finnerty, J.E., Park, H.Y., 1987. Stock index futures: does the tail wag the dog? Financ. Anal. J. 43, 57–61.
Fleming, J., Ostdiek, B., Whaley, R.E., 1996. Trading costs and the relative rates of price discovery in stock, futures, and option markets. J. Futures Mark. 16, 353–387.
French, K.R., Roll, R., 1986. Stock return variances: the arrival of information and the reaction of traders. J. Financ. Econ. 17, 5–26.
Frino, A., Walter, T., West, A., 2000. The lead–lag relationship between equities and stock index futures markets around information releases. J. Futures Mark. 20, 467–487.
Garcia, D., 2013. Sentiment during recessions. J. Financ. 68, 1267–1300.
Gemmill, G., Thomas, D.C., 2002. Noise trading, costly arbitrage, and asset prices: evidence from closed-end funds. J. Financ. 57, 2571–2594.
Glosten, L.R., 1987. Components of the bid–ask spread and the statistical properties of transaction prices. J. Financ. 42, 1293–1307.
Glosten, L.R., Milgrom, P.R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. J. Financ. Econ. 14, 71–100.
Gonzalo, J., Granger, C., 1995. Estimation of common long-memory components in cointegrated systems. J. Bus. Econ. Stat. 13, 27–35.
Greene, J., Smart, S., 1999. Liquidity provision and noise trading: evidence from the "investment dartboard". Column. J. Financ. 54, 1885–1899.
Harris, L., 1989. A day-end transaction price anomaly. J. Financ. Quant. Anal. 24, 29–45.
Hasbrouck, J., 1995. One security, many markets: determining the contributions to price discovery. J. Financ. 50, 1175–1199.
Hasbrouck, J., 2003. Intraday price formation in U.S. equity index markets. J. Financ. 58, 2375–2400.
Hill, R.C., Griffiths, W.E., Lim, G.C., 2008. Principles of Econometrics. Wiley, Hoboken, NJ.
Huang, D., Jiang, F., Tu, J., Zhou, G., 2015. Investor sentiment aligned: a powerful predictor of stock returns. Rev. Financ. Stud. 28, 791–837.
Huang, R.D., Stoll, H.R., 1997. The components of the bid–ask spread: a general approach. Rev. Financ. Stud. 10, 995–1034.
Jones, C.M., Seguin, P.J., 1997. Transaction costs and price volatility: evidence from commission deregulation. Am. Econ. Rev. 87, 728–737.
Käppi, J., 1997. Pricing of futures contracts on coupon bonds: empirical evidence from Finland. Eur. Financ. Manag. 3, 321–332.
Karlsson, N., Loewenstein, G., Seppi, D., 2009. The ostrich effect: selective attention to information. J. Risk Uncertain. 38, 95–115.
Kawaller, I.G., Koch, P.D., Koch, T.W., 1987. The temporal price relationship between S&P 500 futures and the S&P 500 index. J. Financ. 42, 1309–1329.

Kurov, A., 2008. Investor sentiment, trading behavior and informational efficiency in index futures markets. Financ. Rev. 43, 107–127.
Kurov, A., 2010. Investor sentiment and the stock market's reaction to monetary policy. J. Bank. Financ. 34, 139–149.
Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica 53, 1315–1335.
Lee, C.M.C., Mucklow, B., Ready, M.J., 1993. Spreads, depths, and the impact of earnings information: an intraday analysis. Rev. Financ. Stud. 6, 345–374.
Lee, W.Y., Jiang, C.X., Indro, D.C., 2002. Stock market volatility, excess returns, and the role of investor sentiment. J. Bank. Financ. 26, 2277–2299.
Lien, D., Wang, Z., 2016. Estimation of market information shares: a comparison. J. Futures Mark. 36, 1108–1124.
Liu, S., 2015. Investor sentiment and stock market liquidity. J. Behav. Financ. 16, 51–67.
Mayhew, S., Sarin, A., Shastri, K., 1995. The allocation of informed trading across related markets: an analysis of the impact of changes in equity-option margin requirements. J. Financ. 50, 1635–1653.
McInish, T.H., Wood, R.A., 1992. An analysis of intraday patterns in bid/ask spreads for NYSE stocks. J. Financ. 47, 753–764.
Ng, N., 1987. Detecting spot prices forecasts in futures prices using causality tests. Rev. Futures Mark. 6, 250–267.
Roll, R., 1984. A simple implicit measure of the effective bid–ask spread in an efficient market. J. Financ. 39, 1127–1139.
Schmeling, M., 2009. Investor sentiment and stock returns: some international evidence. J. Empir. Financ. 16, 394–408.
Sharpe, W., Alexander, G., 1990. Investment. Prentice Hall, Englewood Cliffs, NJ.
Shleifer, A., Vishny, R.W., 1997. The limits of arbitrage. J. Financ. 52, 35–55.
Smales, L.A., 2014. News sentiment in the gold futures market. J. Bank. Financ. 49, 275–286.
Stambaugh, R.F., Yu, J., Yuan, Y., 2012. The short of it: investor sentiment and anomalies. J. Financ. Econ. 104, 288–302.
Stoll, H.R., 1978. The supply of dealer services in securities markets. J. Financ. 33, 1133–1151.
Stoll, H.R., Whaley, R.E., 1990. The dynamics of stock index and stock index futures returns. J. Financ. Quant. Anal. 25, 441–468.
Tetlock, P.C., 2007. Giving content to investor sentiment: the role of media in the stock market. J. Financ. 62, 1139–1168.
Wang, G.H.K., Michalski, R.J., Jordan, J.V., Moriarty, E.V., 1994. An intraday analysis of bid–ask spreads and price volatility in the S&P 500 index futures market. J. Futures Mark. 14, 837–859.
Wang, G.H.K., Yau, J., 2000. Trading volume, bid–ask spread, and price volatility in futures markets. J. Futures Mark. 20, 943–970.
Yu, J., Yuan, Y., 2011. Investor sentiment and the mean–variance relation. J. Financ. Econ. 100, 367–381.
Yuan, Y., 2015. Market-wide attention, trading, and stock returns. J. Financ. Econ. 116 (3), 548–564. doi:10.1016/j.jfineco.2015.03.006.