Rev Deriv Res
DOI 10.1007/s11147-006-9004-0

Price discovery in the U.S. stock and stock options markets: A portfolio approach

Richard Holowczak · Yusif E. Simaan · Liuren Wu

© Springer Science + Business Media, LLC 2007

Abstract Option prices vary with not only the underlying asset price, but also volatilities and higher moments. In this paper, we use a portfolio of options to seclude the value change of the portfolio from the impact of volatility and higher moments. We apply this portfolio approach to the price discovery analysis in the U.S. stock and stock options markets. We find that the price discovery on the directional movement of the stock price mainly occurs in the stock market, more so now than before as an increasing proportion of options market makers adopt automated quoting algorithms. Nevertheless, the options market becomes more informative during periods of significant options trading activities. The informativeness of the options quotes increases further when the options trading activity generates net sell or buy pressure on the underlying stock price, even more so when the pressure is consistent with deviations between the stock and the options market quotes.

Keywords Price discovery · Options · Stocks · Put-call parity · Automated quoting · Options trade

JEL Classification C52, G10, G13, G14

R. Holowczak
Zicklin School of Business, Baruch College, One Bernard Baruch Way, Box H125,
New York, NY 10010
e-mail: richard_holowczak@baruch.cuny.edu

Y. E. Simaan
Graduate School of Business, Fordham University, 113 West 60th Street,
New York, NY 10023
e-mail: simaan@attglobal.net

L. Wu (✉)
Zicklin School of Business, Baruch College, One Bernard Baruch Way, Box B10-225,
New York, NY 10010
e-mail: liuren_wu@baruch.cuny.edu

In an efficient market, information should impact different markets simultaneously; yet, due to market frictions, different markets may assimilate the same information at different speeds. Price discoveries among different markets have attracted great attentions from both academics and practitioners. Academics test the evidence against market efficiency and various microstructure theories. Practitioners strive to identify profitable trading opportunities and the direction of the information flow.

One of the intriguing areas for price discovery research is the information flow between the market of a primary asset and the markets on its derivatives, such as options. Black (1975) suggested that the higher leverage available in the options markets could induce informed traders to transact options rather than stocks. Easley, O'Hara, and Srinivas (1998) argue that informed traders prefer the options market because they can hide themselves better among uninformed traders due to the availability of multiple options contracts underlied by each single stock. Both arguments imply that price discovery should occur in the options market.

However, the empirical evidence is far from conclusive. Early studies, e.g., Manaster and Rendleman Jr. (1982) and Bhattacharya (1987) find that options market leads the underlying asset market in revealing information, but more recent studies, e.g., Stephan and Whaley (1990), Chan, Chung, and Johnson (1993), Easley, O'Hara, and Srinivas (1998), Jarnecic (1999), Chan, Chung, and Fong (2002) and Chakravarty, Gulen, and Mayhew (2004), have essentially reversed the conclusion.

Underlying the conflicting evidence is the inherent difficulty in assessing the price discovery between the stock and the stock options markets. Unlike quotes of the same asset at different exchanges or comparisons between the spot and the futures markets, the terminal payoff function of an option contract is nonlinear in the underlying stock price. This nonlinear payoff structure dictates that the value change in the option depends on not only the change in the underlying stock price level, but also changes in the variance and higher moments of the underlying stock return.

Variance and higher moments are not directly observable. Therefore, it is difficult to control and separate their impacts when performing price discovery analysis on the directional movement of the underlying stock price. Most studies use some lagged version of implied or historic volatility as an input and use some variants of the Black and Scholes (1973) model to infer the stock price from the option price. In doing so, these studies assume that this volatility input does not change during the next time interval. Given the predominant evidence on stochastic volatility and on the strong correlation between return and volatility,[1] the results from these studies are likely biased by the variations in return volatility and higher moments.

In this paper, we propose a simple procedure to effectively separate the impact of the directional movement of the underlying stock price from the impacts of stochastic volatility and higher moments. Instead of using a single option contract, we form a portfolio with a pair of call and put options at the same strike price and maturity. The

[1] The list of empirical studies on stochastic volatility and its correlation with the stock return is long. The most prominent studies based on time series return data include Black (1976), Andersen, Benzoni, and Lund (2002), Ding, Engle, and Granger (1993), Ding and Granger (1996) and Eraker, Johannes, and Polson (2003). Evidence from the options market includes Bakshi, Cao, and Chen (1997), Bates (1997, 2000), Britten-Jones and Neuberger (2000), Carr and Wu (2003), Eraker (2004), Huang and Wu (2004), Jones (2003), and Pan (2002).

portfolio synthesizes a forward contract on the underlying stock. The terminal payoff to this portfolio is linear in the underlying asset price, even though the payoff to each individual option contract is nonlinear. As a result, the value change of the portfolio depends only on the change in the underlying stock price level, but not on the volatility or any other higher moments of the stock return. Therefore, price discovery analysis between the stock market quotes and the synthetic quotes from this portfolio of options are absent from the biases in earlier studies.

We apply this procedure to the U.S. stock and stock options markets and analyze the information flow between the two markets. We choose 40 stocks that underly the most active option trading activity during our sample period between May and July of 2002. The list of stocks includes 18 from NYSE and 22 from NASDAQ. We compare the quotes from these 40 stocks and the synthetic quotes from the portfolio of the most active options underlying these stocks. Based on the cointegrating relation between the quotes from the two markets, we estimate a vector error-correction model (VECM) and compute the Hasbrouck (1995) information share contributable to the options market. We also compute the loading of the options market on the common factor following the procedure in Gonzalo and Granger (1995).

We find that the information share and common loading estimates are statistically significant on both markets for the 40 stocks, but that the price discovery about the future stock price movement occurs more in the stock market than in the options market, more so for NASDAQ stocks than for NYSE stocks.

One possible explanation for the low information content in the options market is that transaction costs on options are much higher than that on stocks. As long as transaction costs overweigh the benefits of trading with leverage (Black, 1975) and hiding among multiple contracts (Easley, O'Hara, and Srinivas, 1998), the stock market remains the more efficient market for betting on the directional movements in the stock price. Nevertheless, the benefits will increase and can overweigh the transaction cost for investors with private information that points to large price movements. Thus, we hypothesize that the informativeness of options quotes increases in the presence of large stock price movements. To test this hypothesis, we specify a conditional VECM model and define the conditioning variable based on the relative magnitude of price movement. Consistent with our hypothesis, the parameter estimates on the conditional VECM show that the options market becomes significantly more informative when the underlying stock price experiences large moves.

During the past decade, the options market has experienced tremendous growth and consolidation (Battalio, Hatch, and Jennings, 2004; Mayhew, 2002; Simaan and Wu, 2002). Competition is stronger than ever. Options trading volume in 2006 more than doubles that in three years ago. Although the stock market has also been transforming itself, the growth in the options market is definitely more explosive during the past few years than that in the stock market. Therefore, if transaction cost is the main culprit for the low information content in the options market, we would expect the information content to increase over time as the options market becomes more efficient. However, when we repeat the analysis using a more recent sample in March 2006, the information share estimates on the options market become even lower, indicating that the changing market conditions have made the options market even less informative about the directional movement of the underlying stock price. Our information share estimates on the options market are also lower than those reported in Chakravarty, Gulen, and Mayhew (2004), who use an earlier sample between 1988 to 1992. Thus, transaction cost does not represent the full story.

One possible explanation for the declining information content in the options market is related to the recent microstructural transformation in options market making. Despite the rapid and continuing growth in the options market, the options trading activities are much thinner than that in the underlying stock. Yet, whenever the stock price moves in the stock market, the options market makers are forced to update all the quotes on the hundreds of options on the stock, even if there are no options trading activity. This phenomenon leads to dramatically higher quote-to-trade ratios in the options market than in the underlying stock market. To alleviate the uneven quoting burden relative to the sparse trading activity, increasingly many options market making firms have developed automated quote updating algorithms that allow them to update the options quotes automatically whenever the stock price moves in the stock market. In the absence of options order flows, these automated updating algorithms imply that the information flow becomes almost purely one directional from the stock market to the options market. Only when there are significant options transactions do the options quotes adjust accordingly and spill the order flow information over to the stock market. To test this conjecture, we also design a series of conditioning variables related to the trading activities and estimate conditional VECM models with these conditioning variables. The estimation results show that options quotes become much more informative during periods of significant options trading activities. The informativeness of options quotes increases further when the options trading activity generates net sell or buy pressure on the underlying stock price, even more so when the pressure is also consistent with deviations between the stock and the options market quotes.

In related literature, Easley, O'Hara, and Srinivas (1998) aggregate the buy trades on put options with the sell trades on call options, regarding the two types of trades as having the same directional impact on the underlying price movement. This aggregation can be regarded as a trading-volume analogue to our portfolio approach. Also related to our study, Chakravarty, Gulen, and Mayhew (2004) adopt the same Hasbrouck (1995)'s information share measure in analyzing the price discovery between the stock and stock options market. However, they use only quotes from call options and infer the stock price based on a binomial tree model, assuming constant volatility. They use a lagged implied volatility (over 30 minutes) as the volatility input to infer the stock price. Thus, the inferred stock price lumps the variation in the underlying stock price and the variation in the volatility over the past 30 minutes. By forming a portfolio of options, we effectively separate the two types of variations. Furthermore, they base their analysis on option quotes from one exchange, the Chicago Board of Options Exchange (CBOE), from 1988 to 1992. We base our analysis on the national best bid offer (NBBO) on multiple-exchange listed options in 2002 and 2006. The different sample periods allow us to analyze how the information flow has changed over time with varying market conditions.

Most recently, Pan and Poteshman (2006) obtain a unique data set that decomposes the daily trading volume on each option into 16 categories defined by four trade types and four investor classes. They construct put-call ratios from option volume initiated by buyers to open new positions, and find that stocks with low put-call ratios outperform stocks with high put-call ratios by more than 40 basis points on the next day. Their finding is consistent with our conditional VECM results on options trading activities.

Furthermore, their result also shows that the options trading activity can become even more informative when one can distinguish the trading types.

The remainder of this paper is organized as follows. Section 1 describes the data source, the summary statistics of the selected sample, and the econometrics underlying the results of this paper. Section 2 discusses the results. Section 3 discuss the findings based on a more recent sample in 2006. Section 4 concludes.

1 Sample selection and econometric specification

Our data sources are the electronic message feeds from S&P ComStock XpressFeed. The message feeds contain real-time updates on quotes and transactions of U.S. stocks and stock options. The feeds include quote price, quote size, transaction price, and transaction size at each exchange.

The options data provider is the Options Price Reporting Authority (OPRA). In 1981, the Securities and Exchange Commission approved the OPRA plan as a national market system plan, pursuant to Sections 11A(a)(2) and 11A(a)(3)(B) of the Act. The OPRA plan governs the process by which options market data are collected, consolidated, and disseminated.

As of 2002, during which we draw our first sample, stock options trade on five options exchanges in the United States: The American Stock Exchange (AMEX), the Chicago Board of Options Exchange (CBOE), the International Securities Exchange (ISE), the Pacific Stock Exchange (PCX),[2] and the Philadelphia Stock Exchange (PHLX). On February 6, 2004, yet another options exchange, Boston Options Exchange (BOX), joined the competition in stock options trading. OPRA receives the information on each transaction and each quote update from these exchanges as a coded "message." Trade and quote data generate continuously during the hours that markets are open for each options product listed on each options exchange. Beginning April 1st, 2002, in addition to the local quotes from each exchange, OPRA also reports a national best bid offer (NBBO) for each quote.

1.1 Sample selection

To build our sample, we first extract all the option trades in May, 2000 from the OPRA message. Then we rank the options contracts in terms of trading activity as measured by the number of trades. On the list, we only retain options contracts that are traded on all five options exchanges. We then search through the list to identify the top 40 stocks that underly the most active options contracts. Table 1 provides a full list of the 40 stocks. It includes 18 NYSE stocks, 21 NASDAQ stocks, and QQQ, the NASDAQ 100 Trust fund. Although QQQ is listed at the NYSE, we classify it as a NASDAQ stock when we report averages across stocks at different exchanges.

We extract stocks and stock options from the message feeds based on the list of 40 stocks over 52 business days from May 6th, 2002 to July 19th, 2002. Table 2 presents the summary statistics on the option trading activities underlying these 40 stocks over

[2] PCX is now owned by NYSE Arca Options, a new platform for trading options from the NYSE Group.

Table 1 Stock list

TICKER	COMPANY NAME	TICKER	COMPANY NAME
Index:		*NASDAQ Stocks:*	
QQQ	NASDAQ 100 TRUST	SUNW	SUN MICROSYSTEMS
		MSFT	MICROSOFT
NYSE Stocks:		CSCO	CISCO SYSTEMS
TYC	TYCO INTL	INTC	INTEL CORP
GE	GENERAL ELECTRIC CO	NVDA	NVIDIA CORP
AOL	AOL TIME WARNER	AMAT	APPLIED MATERIAL
IBM	INTL BUS MACHINE	ORCL	ORACLE CORP
JPM	JP MORGAN CHASE	QCOM	QUAL COMM INC
EMC	EMC CORP	DELL	DELL COMPUTER
WMT	WAL-MART STORES	AMGN	AMGN
HD	HOME DEPOT INC	JNPR	JUNIPER NETWORKS
C	CITIGROUP	SEBL	SIEBEL SYSTEMS
VZ	VERIZON COMMS	QLGC	QLOGIC CORP
NEM	NEWMONT MINING	VRTS	VERITAS SOFTW
CPN	CALPINE CORP	EBAY	EBAY INC
MO	PHILIP MORRIS	BRCM	BROADCOM CORP
JNJ	JOHNSON&JOHNSON	XLNX	XILINIX INC
XOM	EXXON MOBIL	KLAC	KLA TENCOR
HAL	HALLIBURTON CO	BRCD	BROADE COMMS
PFE	PFIZER INC	NVLS	NOVELLUS SYS
BA	BOEING CO	GNSS	GENESIS MICRO

Entries report the list of stocks that are used for the empirical analysis in this paper.

the 52 business days. For each stock, we report the total number of option transactions, the aggregate trading volume, the standard deviation, minimum, and maximum of the trading volume for transactions on both calls (left) and puts (right). QQQ towers all stocks in underlying the most actively traded options. TYC underlies the most actively traded options on single name stocks during our sample period. The most active single name stock on NASDAQ in terms of options trading during our sample period is MSFT.

For each stock and at each day, we first extract the NBBO on the underlying stock. We then identify the most active option contract on this stock in terms of number of trades. We also identify the call-put counterpart of this most active option contract. We record the NBBO updates on the stock and the identified options contracts, and expand the NBBO updates into a second-by-second quote book. For multiple quotes that are recorded on the same second, we take the last of these updates at that second.

We analyze the interactions between the stock and options markets based on the midquotes of the national best bid and offer from the two markets. In computing the midquotes, we explicitly account for the potential differences in the bid and ask sizes. Let AP, AS, BP, BS denote the ask price, ask size, bid price, bid size, respectively, we compute the midquote as,

$$MP = \frac{AS \times BP + BS \times AP}{AS + BS}, \tag{1}$$

where weighting is such that the midquote MP tilts toward the quote with a smaller quote size. Although more trading activity often conveys more information,

Table 2 Summary statistics of option trades

TICKER	N		SUM		STD		MIN		MAX	
QQQ	138,662	129,411	9,307,033	7,402,539	353.1	321.1	1	1	18,000	11,000
TYC	45,710	27,032	1,864,123	1,068,913	184.1	168.0	1	1	6,600	8,000
GE	19,004	11,495	521,264	394,151	184.3	150.6	1	1	12,000	6,000
AOL	13,750	8,607	536,385	360,651	222.2	161.0	1	1	8,000	6,250
IBM	23,199	23,608	514,957	616,613	98.1	111.2	1	1	5,665	5,500
JPM	9,307	11,177	291,875	348,037	195.5	156.0	1	1	7,500	5,000
EMC	10,157	4,199	451,060	217,302	271.5	333.4	1	1	9,000	7,900
WMT	8,457	4,759	173,205	122,801	119.1	86.6	1	1	8,541	2,300
HD	10,170	3,714	198,814	89,775	146.6	87.2	1	1	8,400	2,000
C	11,402	6,541	324,954	306,769	204.2	225.6	1	1	16,000	7,000
VZ	4,741	8,130	116,177	163,396	137.9	106.7	1	1	5,000	5,000
NEM	6,068	2,356	103,020	52,603	41.0	59.0	1	1	1,000	1,000
CPN	5,291	3,362	138,855	113,383	125.2	169.6	1	1	4,000	4,300
MO	5,851	3,311	191,782	267,478	119.1	300.0	1	1	3,000	4,499
JNJ	11,846	6,700	264,610	248,539	95.5	167.8	1	1	3,504	6,200
XOM	7,718	3,579	172,463	68,046	130.2	79.2	1	1	6,000	1,590
HAL	4,116	2,384	123,033	53,849	135.6	95.1	1	1	3,500	3,250
PFE	8,736	4,525	256,171	147,059	97.1	120.1	1	1	2,500	2,500
BA	3,138	2,689	55,396	44,258	78.0	50.7	1	1	2,001	1,000
SUNW	14,698	5,431	519,232	225,323	158.6	355.2	1	1	6,500	17,900
MSFT	27,486	18,263	904,267	746,126	168.3	352.0	1	1	9,150	20,035
CSCO	19,305	10,431	680,619	570,606	165.5	282.6	1	1	9,999	10,000
INTC	25,829	14,029	877,006	585,647	233.6	238.8	1	1	27,000	13,500
NVDA	11,744	7,894	172,007	124,368	45.0	52.8	1	1	1,816	2,500
AMAT	16,884	10,525	412,227	364,105	108.6	167.5	1	1	5,000	5,000
ORCL	9,936	4,351	334,872	161,118	159.7	175.7	1	1	7,000	9,002
QCOM	12,500	9,143	220,008	171,383	68.2	74.6	1	1	3,350	1,901
DELL	12,254	8,048	371,663	400,062	142.3	209.9	1	1	5,000	5,900
AMGN	14,587	7,234	290,072	147,224	73.2	100.1	1	1	5,000	5,000
JNPR	8,168	3,969	135,702	147,281	35.7	297.5	1	1	950	9,000
SEBL	14,418	10,004	364,741	314,385	133.0	165.9	1	1	7,630	10,000
QLGC	13,500	13,644	211,579	215,643	44.5	38.4	1	1	1,500	1,300
VRTS	6,517	6,052	163,862	166,938	129.8	177.2	1	1	5,000	5,000
EBAY	13,015	11,101	263,055	261,644	88.5	75.0	1	1	5,259	2,000
BRCM	8,719	7,683	156,435	166,850	44.3	106.6	1	1	1,000	4,000
XLNX	6,488	6,650	135,651	171,197	70.5	128.0	1	1	3,000	4,500
KLAC	7,191	6,779	142,899	148,726	71.5	112.0	1	1	2,000	3,500
BRCD	8,546	6,434	160,579	139,190	66.4	80.8	1	1	4,000	4,000
NVLS	5,269	4,163	99,453	109,757	56.5	113.8	1	1	1,500	3,000
GNSS	4,851	2,694	76,016	58,659	70.8	177.1	1	1	2,850	8,000

Entries report the summary statistics of trades on all call options and put options underlying each stock: Number of trades (N), aggregate trading volume (SUM), standard deviation (STD), minimum (MIN), and maximum (MAX) of the trading volume. Under each property, the left column denotes call options and the right column put options. The data are from May 6th, 2002 to July 19th, 2002, a span of 52 business days.

a quote with a smaller quote size is closer to the mid-quote than quotes with larger quote sizes. A market maker can afford to be more aggressive in her bid-ask quotes when she only honors a small quantity and hence subjects herself to lesser risk.

1.2 Inferring the stock price from a portfolio of options

In analyzing the dynamic interactions between the stock market quotes and the quotes from the options market, an important concern is that the price of options depends not only on the underlying stock price level, but also on the stock return volatility and higher moments. This multi-dimensional dependence is a direct result of the nonlinear payoff structure of an option. For example, when the stock return volatility increases, the values of both call and put options increase, even if the underlying stock price remains unchanged. Therefore, the issue of how to identify and control variables other than the underlying stock price has plagued previous studies on price discovery between stocks and stock options. Here, we propose a portfolio approach that exploits put-call parity to separate the impact of the underlying stock price level from the impacts of the stock return volatility and higher moments.

Instead of directly using the quotes on a single option, we form a portfolio with a call option and a put option at the same strike and maturity. The terminal payoff to this portfolio is linear in the underlying stock price level, although the payoff to each individual option component is nonlinear. Specifically, we exploit the put-call parity relation between the put (P) and call (C) options at the same strike price (K) and maturity (τ),

$$C_t - P_t = S_t e^{-d_t \tau} - K e^{-r_t \tau}, \tag{2}$$

where d_t and r_t denote the dividend yield and interest rate at time t, respectively. This equality holds exactly for European options. For the American options that we are dealing with, we use this relation as an approximation in forming a portfolio of options by longing one call and shorting one put option at the same strike price and maturity level. The linear payoff structure of the portfolio dictates that stock volatility or higher moment no longer exert significant influence on the portfolio value.

We acknowledge that variations in the value of the portfolio can also be induced by movements in the early exercise premium of the American options, the dividend yield, and the interest rate. Nevertheless, given the small intraday variation of these components as compared to the variation of the stock price, we can effectively regard them as constants at each date in identifying the variation in the stock price.

Therefore, at each date and for each stock in our sample, we first identify the most active option in terms of number of trades, as well as its put-call counterpart at the same strike and maturity. Then, we define the following quantity based on the midquote of the NBBO,

$$CP_t = C_t - P_t, \tag{3}$$

which we can regard as the midquote of the portfolio.

Since we perform the price discovery analysis on different days in one integrated analysis, we need to stack the daily data together. However, the chosen option pairs can be different at different days depending on the option trading activities. Thus, the relative magnitude of the portfolio value CP_t at different dates can differ by approximately a constant. To circumvent this problem, we run the following regression at each date between the stock midquotes and the option portfolio value,

$$S_t = a + b\,CP_t + e_t. \tag{4}$$

We regard the fitted value $O_t \equiv \widehat{a} + \widehat{b}CP_t$ as the synthetic stock price from the options quotes. Then, we perform the price discovery analysis between the pair of stock prices from the two markets, (S_t, O_t).

1.3 Unit root and cointegration

Stock prices are often assumed to have a random walk component. Since S_t and O_t underly the value of the same stock, albeit from different markets, we expect them to be cointegrated. As a preliminary analysis, we first check the order of integration for the two price series S_t and O_t, as well as their first differences. We use the augmented Dickey and Fuller (1979) test with 20 lags. Table 3 reports the p-values of the augmented Dickey-Fuller test on the levels and first differences with the null hypothesis that the series contains a unit root. Thus, a p-value that is close to zero implies that the null hypothesis of unit root is rejected. We perform all the tests on the logarithms of the price quote. The tests on the first differences of all series generate p-values less than 0.005, strongly rejecting the null hypothesis of a unit root in the differences and implying that the order of integration in the price level is at most one. However, for most of the stocks, the test cannot reject the null hypothesis that there is a unit root in the logarithm of the price level. A few exceptions occur for this test. For the midquotes from the stock market during our sample period, the null hypothesis of a unit root in the price level is rejected for TYC, WMT, MSFT, and DEL at 5% critical value.

As a preliminary test for cointegration, we run the simple regression,

$$s_t = a + bo_t + R_t,$$

and apply the unit root test on the regression residual R_t. In the regression, s_t and o_t denote the logarithms of the corresponding quotes S_t and O_t, respectively. We report the test results in Table 3 under the column title "R_t." The null hypothesis of a unit root in either the residual level or the residual difference is strongly rejected for all stocks, confirming our conjecture that the quotes from the stock market and the quotes from the option portfolio are cointegrated.

The columns in Table 3 under the title "Regression Coefficients" report the intercept and slope estimates of the above regression, with the p-values of the estimates reported in the parentheses. The intercept is indistinguishable from zero in most cases, and the slope estimates are all at one. This uniform estimates are a result of our rescaling of the option implied stock price. Therefore, under this rescaling, we have a cointegrating relation of $[1, -1]$ for the two sources of stock prices.

Johansen (1988, 1991) proposes a formal test of cointegration based on a vector error correction model (VECM). If we use $y_t \equiv [s_t, o_t]^\top$ to denote the log price vector,

Table 3 Unit root and cointegration tests

	Dickey-fuller tests						Regression coefficients				$\lambda_{\max}$	
	s_t		o_t		R_t		$\widehat{a}$		$\widehat{b}$			
TICKER	Level	Diff.	Level	Diff.	Level	Diff.	Est.	p-value	Est.	p-value	0	1
QQQ	0.86	0.00	0.91	0.00	0.00	0.00	−0.00	(0.96)	1.00	(0.00)	1992.28	0.52
TYC	0.04	0.00	0.63	0.00	0.00	0.00	0.00	(0.39)	1.00	(0.00)	17659.74	1.63
GE	0.73	0.00	0.91	0.00	0.00	0.00	−0.00	(0.53)	1.00	(0.00)	2292.53	0.17
AOL	0.91	0.00	0.99	0.00	0.00	0.00	−0.00	(0.05)	1.00	(0.00)	1616.71	0.21
IBM	0.69	0.00	0.80	0.00	0.00	0.00	−0.00	(0.45)	1.00	(0.00)	5143.89	1.44
JPM	0.07	0.00	0.99	0.00	0.00	0.00	0.00	(0.00)	1.00	(0.00)	222461.36	0.30
EMC	0.09	0.00	0.37	0.00	0.00	0.00	0.00	(0.65)	1.00	(0.00)	2927.54	3.70
WMT	0.00	0.00	1.00	0.00	0.00	0.00	−0.00	(0.49)	1.00	(0.00)	2155.39	0.45
HD	0.96	0.00	0.99	0.00	0.00	0.00	−0.00	(0.49)	1.00	(0.00)	2950.89	0.45
C	0.75	0.00	0.89	0.00	0.00	0.00	−0.00	(0.43)	1.00	(0.00)	2959.03	0.40
VZ	0.95	0.00	0.98	0.00	0.00	0.00	−0.00	(0.59)	1.00	(0.00)	2530.75	0.02
NEM	0.48	0.00	0.79	0.00	0.00	0.00	−0.00	(0.79)	1.00	(0.00)	1816.44	1.50
CPN	0.95	0.00	0.98	0.00	0.00	0.00	−0.00	(0.00)	1.00	(0.00)	1387.76	0.13
MO	0.99	0.00	1.00	0.00	0.00	0.00	−0.00	(0.50)	1.00	(0.00)	1629.30	1.24
JNJ	0.99	0.00	1.00	0.00	0.00	0.00	−0.00	(0.32)	1.00	(0.00)	2650.47	1.19
XOM	1.00	0.00	1.00	0.00	0.00	0.00	0.01	(0.00)	1.00	(0.00)	3507.88	3.61
HAL	1.00	0.00	1.00	0.00	0.00	0.00	−0.00	(0.18)	1.00	(0.00)	1323.18	4.84
PFE	0.97	0.00	0.98	0.00	0.00	0.00	−0.00	(0.76)	1.00	(0.00)	4010.72	0.01
BA	0.11	0.00	0.49	0.00	0.00	0.00	−0.00	(0.42)	1.00	(0.00)	2668.81	3.16
SUNW	0.87	0.00	0.90	0.00	0.00	0.00	−0.00	(0.07)	1.00	(0.00)	1144.94	1.43
MSFT	0.03	0.00	0.07	0.00	0.00	0.00	−0.00	(0.82)	1.00	(0.00)	1887.47	5.55
CSCO	0.25	0.00	0.38	0.00	0.00	0.00	−0.00	(0.29)	1.00	(0.00)	923.74	6.03
INTC	0.81	0.00	0.83	0.00	0.00	0.00	−0.00	(0.14)	1.00	(0.00)	1059.31	0.72
NVDA	0.92	0.00	0.95	0.00	0.00	0.00	−0.00	(0.19)	1.00	(0.00)	2767.15	0.00
AMAT	0.78	0.00	0.85	0.00	0.00	0.00	−0.00	(0.24)	1.00	(0.00)	1442.56	0.55
ORCL	0.14	0.00	0.25	0.00	0.00	0.00	−0.00	(0.97)	1.00	(0.00)	1238.15	3.27
QCOM	0.12	0.00	0.21	0.00	0.00	0.00	−0.00	(0.25)	1.00	(0.00)	1987.46	4.06
DELL	0.03	0.00	0.10	0.00	0.00	0.00	−0.00	(0.09)	1.00	(0.00)	1261.38	6.88
AMGN	0.80	0.00	0.86	0.00	0.00	0.00	−0.00	(0.15)	1.00	(0.00)	2429.29	0.58
JNPR	0.68	0.00	0.67	0.00	0.00	0.00	−0.00	(0.00)	1.00	(0.00)	902.04	0.85
SEBL	0.97	0.00	0.98	0.00	0.00	0.00	−0.00	(0.05)	1.00	(0.00)	2721.09	0.01
QLGC	0.45	0.00	0.55	0.00	0.00	0.00	−0.00	(0.17)	1.00	(0.00)	3208.02	1.87
VRTS	0.56	0.00	0.69	0.00	0.00	0.00	−0.00	(0.08)	1.00	(0.00)	2235.71	1.68
EBAY	0.09	0.00	0.07	0.00	0.00	0.00	0.00	(0.65)	1.00	(0.00)	4003.51	3.20
BRCM	0.31	0.00	0.34	0.00	0.00	0.00	−0.00	(0.34)	1.00	(0.00)	2026.23	2.88
XLNX	0.92	0.00	0.95	0.00	0.00	0.00	−0.00	(0.16)	1.00	(0.00)	2203.91	0.03
KLAC	0.70	0.00	0.74	0.00	0.00	0.00	−0.00	(0.33)	1.00	(0.00)	2294.08	1.71
BRCD	0.21	0.00	0.29	0.00	0.00	0.00	−0.01	(0.00)	1.00	(0.00)	1253.06	3.48
NVLS	0.86	0.00	0.90	0.00	0.00	0.00	−0.00	(0.13)	1.00	(0.00)	1937.00	0.18
GNSS	0.86	0.00	0.89	0.00	0.00	0.00	−0.00	(0.00)	1.00	(0.00)	2554.39	0.15

(*Continued on next page*)

Table 3 (*Continued*)

	Dickey-fuller tests						Regression coefficients				$\lambda_{\max}$	
	s_t		o_t		R_t		$\widehat{a}$		$\widehat{b}$			
TICKER	Level	Diff.	Level	Diff.	Level	Diff.	Est.	p-value	Est.	p-value	0	1
NYSE	0.65	0.00	0.88	0.00	0.00	0.00	0.00	(0.39)	1.00	(0.00)	15649.58	1.36
NASD	0.55	0.00	0.61	0.00	0.00	0.00	−0.00	(0.28)	1.00	(0.00)	1976.03	2.07
ALL	0.60	0.00	0.73	0.00	0.00	0.00	−0.00	(0.33)	1.00	(0.00)	8129.13	1.75

Columns under "Dickey-Fuller Tests" report the p-values of the augmented Dicky-Fuller tests on the level and first difference of the second by second price series. s_t denotes the logarithm of the stock price from the stock market, o_t denotes the logarithm of the stock price implied from the options market, R_t denotes residuals from the following simple regression,

$$s_t = a + bo_t + R_t .$$

Under each series, the left column (Level) reports the test results on the level and the right panel (Diff.) reports the test results on the first difference. A p-value of close to zero implies that the null of a unit root is rejected. The intercept and slope estimates (p-values in the parenthesis) of the regression are reported in columns under "Regression Coefficients." The last two columns report the Johansen's maximum eigenvalue test of the null hypothesis that there are exactly zero (left column) or one (right column) cointegrating relation. The critical value at one percent level is 25.75 under the null of zero cointegration and 19.19 under the null of one cointegration.

we can write the VECM as,

$$\Delta y_t = \mu + A_0 y_{t-1} + A_j \sum_{j=2}^{k} \Delta y_{t-j} + \varepsilon_t , \tag{5}$$

where A_j, $j = 0, 1, \ldots , k$, are $(n \times n)$ coefficient matrices, with $n = 2$ in our application. The term μ is a constant $(n \times 1)$ vector, and ε_t denotes the vector regression residual. Johansen (1988, 1991) shows that the rank of the coefficient matrix A_0 represents the number of cointegration relationships present among the variables y_t. Furthermore, suppose that A_0 has a reduced rank of $r < n$, we can perform the decomposition $A_0 = \alpha\beta^\top$, where α and β are both $(n \times r)$ matrices. The β matrix consists of the cointegration vectors, and α is the error correction (or equilibrium adjustment) matrix.

Johansen (1988, 1991) proposes a maximum likelihood test on the rank, based on the following procedure. Let $\mathcal{S} = (\Sigma_{u\,u})^{-1}\, \Sigma_{u\,v}\, (\Sigma_{v\,v})^{-1}\, \Sigma_{v\,u}$, where Σ_{xy} denotes the covariance matrix of x and y. The subscript u and v denote the regression residuals based on levels and first differences, respectively. Let $\lambda_1 > \lambda_2 > \cdots > \lambda_n$ denote the eigenvalues of $\mathcal{S}$ and let $M = (m_1, m_2, \ldots , m_n)$ denote the matrix formed by the corresponding eigenvectors, with the normalization such that $M^\top \Sigma_{u\,u} M = I$. Then, Johansen's maximum eigenvalue test of the null hypothesis that there are r cointegration relations against $r + 1$ is,

$$\lambda_{\max} = -T \ln (1 - \lambda_{r+1}) .$$

As λ_{r+1} approaches zero, $\lambda_{\max}$ approaches zero as well.

We perform the test on the 40 stocks in our sample. In running the vector regressions, we include lagged values up to 60 seconds, but experiments with other lag lengths and specifications (such as the moving average and polynomial-distributed lags proposed in Hasbrouck (2003)) generate the same qualitative results. The last two columns of Table 3 report the estimates of $\lambda_{\max}$ for $r = 0$ and $r = 1$, respectively. The results on all 40 stocks unanimously state that the two series s_t and o_t are cointegrated. The estimates of $\lambda_{\max}$ under the null hypothesis of no cointegration ($r = 0$) range between 902.04 and 222461.36 for the 40 stocks, with a sample average at 8129.13. Based on the critical values reported in Osterward-Lenum (1992) (25.75 at 1% level), the null hypothesis of no cointegration is strongly rejected for all stocks. In contrast, the estimates of $\lambda_{\max}$ under the null hypothesis of one cointegration ($r = 1$) range between 0.00 and 6.88, with a sample average of 1.75. Since the critical value at 1% level is 19.19, the null hypothesis of one cointegration cannot be rejected. In short, different tests concur that the stock quotes and the quotes of the options portfolio are cointegrated.

1.4 The vector error correction model

Given the strong evidence on cointegration and the cointegrating relation of $[1, -1]$ as a result of our rescaling, we can rewrite the vector error correction model (VECM) as,

$$\Delta y_t = A_1 \Delta y_{t-1} + A_2 \Delta y_{t-2} + \cdots + A_k \Delta y_{t-k} + \gamma (s_{t-1} - o_{t-1}) + u_t, \qquad (6)$$

where the coefficient matrices A_i are (2×2) square matrices, γ is a 2×1 vector, and u_t denotes the innovation vector. The term $(s_{t-1} - o_{t-1})$ is formed based on the $[1, -1]$ cointegrating relation and represents an error correction term due to the cointegration. The coefficients A_i capture how previous price changes (log returns) impact the future price changes. The coefficient γ further "corrects" the forecasts based on how much the quotes from the two markets are deviating from their long-run cointegrating relation. In estimating the VECM model, we include lagged second-by-second returns up to 60 seconds. The results are qualitatively the same with different lag specifications.

Based on the VECM model in Eq. (6), Hasbrouck (1995) proposes an information share (*IS*) that measures the relative contribution in price discovery from the two markets. Hasbrouck assumes that the "efficient price" of the stock follows a random walk and represents the permanent component. Then, the information share decomposes the variance of this efficient price changes into components attributable to the two markets. To compute the information share, we first write out the moving average representation of the VECM model,

$$\Delta y_t = B_0 u_t + B_1 u_{t-1} + B_2 u_{t-2} + \cdots, \quad B_0 = I. \qquad (7)$$

We can calculate the moving average coefficients B by "forecasting" the system subsequent to a unit perturbation. The response to the permanent component is obtained by taking the limit,

$$\mathcal{B} = \lim_{k\to\infty} \sum_{i=0}^{k} B_k. \tag{8}$$

The two rows of $\mathcal{B}$ are identical. Let b be either row of $\mathcal{B}$, and let Ω denote the covariance matrix of innovation vector u, the variance of the common random walk component of the quotes is,

$$\sigma_w^2 = b\Omega b^\top.$$

If Ω is diagonal, the information share of the i-th market is defined as,

$$IS_i = \frac{b_i^2 \Omega_{ii}^2}{\sigma_w^2}. \tag{9}$$

When Ω is not diagonal, the information share is not uniquely defined. Instead, we can obtain the lower and upper bounds by considering the Cholesky factorization of all the rotations of the disturbances.

As an alternative price discovery measure, Harris, McInish, and Wood (2002b) follows Gonzalo and Granger (1995) in identifying a permanent component with two properties: (1) This permanent component is a linear combination of the contemporaneous quotes from the two markets, and (2) the permanent component is not Granger-caused in the long run by the stationary linear combinations of the quotes. Given the estimates for the error correction coefficients $\gamma \equiv [\gamma_1, \gamma_2]^\top$, the normalized loading ($w = [w_1, w_2]^\top$) of this common factor on the quotes from the two markets is given by,

$$\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \frac{1}{\gamma_1 - \gamma_2} \begin{pmatrix} -\gamma_2 \\ \gamma_1 \end{pmatrix}. \tag{10}$$

The common factor identified from this approach is,

$$f_t = w_1 s_t + w_2 o_t.$$

The Journal of Financial Markets devotes most of its July 2002 issue to the interpretation and comparison of the two measures.[3] In essence, the common factor loading defined in Eq. (10) measures the long-run impact of the current price innovations on future prices, whereas the Hasbrouck information share measures the relative contribution of each price series to a random walk component. An integrated analysis of the

[3] The major review and comparison studies in that issue include de Jong (2002), Harris, McInish, and Wood (2002a), Hasbrouck (2002) and Lehmann (2002). Other examples of microstructure applications of the Hasbrouck information share include Hasbrouck (2003) and Chakravarty, Gulen, and Mayhew (2004). Examples of the common factor loading approach in the market microstructure literature include Baillie, Booth, Tse, and Zabotina (2002), Booth, So, and Tse (1999), Chu et al. (1999) and Harris, McInish, and Wood (2002b).

information share, the common factor loading, and the variance covariance structure of the price innovations is useful to gain a full picture of the price discovery between the two markets.

2 Price discovery in the options market

Based on the estimated VECM model on each stock, this section analyzes the results on the information flow between the stock market and the options market.

2.1 Common factor loadings and the information share

Table 4 repots the estimates on the loading of the options market on the common factor (w_2) and the lower and upper bounds of the information share attributable to the options market (IS_2). The information share attributable to the options market is low for most stocks. The average of the information share for the 18 NYSE stocks is 12.7% and 14% for the lower and upper bounds, respectively. The average lower and upper bounds for the 22 NASDAQ stocks are 3.1% and 15.3%, respectively.

Table 4 Common factor loading and information shares attributable to the options market

TICKER	w_2	IS_2		TICKER	w_2	IS_2	
Index:				*NASDAQ Stocks:*			
QQQ	0.249	0.069	0.279	SUNW	0.147	0.023	0.380
				MSFT	0.047	0.002	0.072
NYSE Stocks:				CSCO	0.143	0.031	0.065
TYC	0.843	0.456	0.463	INTC	0.161	0.023	0.028
GE	0.246	0.034	0.054	NVDA	0.199	0.058	0.078
AOL	0.323	0.063	0.067	AMAT	0.155	0.024	0.097
IBM	0.345	0.063	0.121	ORCL	0.109	0.016	0.078
JPM	0.958	0.556	0.572	QCOM	0.198	0.039	0.246
EMC	0.406	0.137	0.140	DELL	0.162	0.024	0.177
WMT	0.277	0.065	0.075	AMGN	0.078	0.006	0.060
HD	0.524	0.261	0.274	JNPR	−0.031	0.001	0.221
C	0.343	0.080	0.082	SEBL	0.120	0.012	0.255
VZ	0.246	0.047	0.052	QLGC	0.176	0.029	0.262
NEM	0.238	0.026	0.029	VRTS	0.248	0.051	0.103
CPN	0.264	0.049	0.051	EBAY	0.197	0.051	0.194
MO	0.385	0.114	0.119	BRCM	0.248	0.049	0.096
JNJ	0.202	0.033	0.090	XLNX	0.196	0.031	0.111
XOM	0.399	0.118	0.126	KLAC	0.158	0.030	0.060
HAL	0.143	0.015	0.016	BRCD	0.158	0.024	0.210
PFE	0.354	0.052	0.059	NVLS	0.205	0.033	0.036
BA	0.425	0.116	0.125	GNSS	0.253	0.054	0.248
AVERAGE	0.384	0.127	0.140	AVERAGE	0.163	0.031	0.153

Entries report the common factor loading (w_2) and information share (IS_2) estimates attributable to the options market for the 40 stocks. Under the information share, the left column reports the lower bound and the right column reports the upper bound. These estimates are computed based on the vector error correction model (VECM) in (6) with second by second lags up to 60 seconds.

Table 5 Covariance matrix of the VECM innovation vector

TICKER	$\sqrt{\Omega_{11}}$	$\sqrt{\Omega_{22}}$	ρ_{12}	TICKER	$\sqrt{\Omega_{11}}$	$\sqrt{\Omega_{22}}$	ρ_{12}
Index:				*NASDAQ Stocks:*			
QQQ	1.715	1.612	0.286	SUNW	3.069	3.505	0.490
				MSFT	1.471	1.543	0.219
NYSE Stocks:				CSCO	1.588	1.747	0.082
TYC	18.715	3.211	0.006	INTC	1.874	1.529	0.015
GE	1.895	1.110	0.047	NVDA	3.473	3.518	0.040
AOL	3.104	1.704	0.006	AMAT	2.554	2.306	0.160
IBM	2.410	1.222	0.102	ORCL	2.354	2.547	0.157
JPM	30.643	1.544	0.015	QCOM	3.048	2.826	0.314
EMC	4.399	2.577	0.004	DELL	2.002	1.827	0.274
WMT	1.439	1.002	0.020	AMGN	2.137	2.062	0.167
HD	2.437	1.328	0.015	JNPR	4.204	4.529	0.494
C	2.066	1.170	0.002	SEBL	3.880	3.592	0.407
VZ	2.080	1.425	0.011	QLGC	2.968	2.788	0.357
NEM	2.017	1.060	0.010	VRTS	3.523	2.566	0.099
CPN	4.477	2.861	0.006	EBAY	2.333	2.396	0.226
MO	1.190	0.688	0.006	BRCM	3.321	2.351	0.092
JNJ	1.512	1.149	0.121	XLNX	3.311	2.548	0.163
XOM	1.470	0.813	0.013	KLAC	2.274	2.184	0.073
HAL	1.880	1.416	0.003	BRCD	3.684	3.498	0.316
PFE	2.581	1.108	0.014	NVLS	2.764	1.990	0.008
BA	1.974	0.975	0.013	GNSS	5.266	4.161	0.284
AVERAGE	4.794	1.465	0.023	AVERAGE	2.855	2.619	0.215

Entries report the covariance matrix of the VECM innovation vector for each stock. Ω_{11} and Ω_{22} denote the two diagonal elements of the covariance matrix (both scaled up by 10^4) and ρ_{12} denotes the correlation of the two elements of the innovation vector. The estimates are computed based on the vector error correction model (VECM) in (6) with second by second lags up to 60 seconds.

On average, the common factor loading estimate on the options market is larger than the corresponding Hasbrouck information share estimates. This "upward bias" can be attributed to the different variance magnitudes from the price innovations of the two markets. Table 5 reports the covariance matrix (Ω) of the price innovation u_t. On average, the variance of the stock market price innovation is larger than the variance of the option market price innovation. Lehmann (2002) notes that one way to link the common factor loading w to the information share is $IS = w\sqrt{\Omega}$, where $\sqrt{\Omega}$ denotes the variance decomposition, $\Omega = \sqrt{\Omega}\sqrt{\Omega}^{\top}$. Since this variance decomposition is not unique when Ω is not diagonal, the indeterminacy of the information share remains. Nevertheless, this relation shows that the larger magnitude of Ω_{11} than Ω_{22} explains why the information share attributable to the options market is smaller than the common factor loading to this market.

Our information share estimates based on the current market conditions differ from the estimates by Chakravarty, Gulen, and Mayhew (2004) based on an earlier sample period from 1988 to 1992. Considering the dramatic improvement in the options market condition during the past decade, we would have expected that our information share estimates be higher than their estimates, but the opposite is observed. For the list of stocks that are considered in both papers, including BA, C, GE, HAL, JNJ, WMT,

Table 6 Average estimates for the error correction coefficients

Model	γ_1		γ_2	
NYSE	−0.0037	(−13.41)	0.0016	35.42
NASD	−0.0003	(−5.27)	0.0015	29.04
GRAND	−0.0018	(−8.93)	0.0016	31.91

Entries report the sample average across the 40 stocks of the estimates and t-values (in parentheses) for the error correction coefficients $\gamma = [\gamma_1, \gamma_2]^\top$ in the vector error correction model in (6).

XON, and MOB,[4] the option market information share estimates in Chakravarty, Gulen, and Mayhew (2004) are uniformly higher than our estimates. We conjecture that the differences can partly be attributed to the different approaches that the two papers use to extract the stock price from the options market. In our paper, we form a portfolio with a pair of put and call options to linearize the payoff function and hence to exclude the impact of volatility and higher moments. In contrast, Chakravarty, Gulen, and Mayhew (2004) use only call options and use a binomial tree approach, with lagged implied volatility as inputs, to infer the current stock price. Beside the methodology differences, the different estimates may also be related to the microstructural transformations in the options market making procedures. We analyze this conjecture in more detail in the next section.

2.2 Parameter estimates of the VECM

To gain further understanding on the dynamic interactions between quotes from the two markets, we analyze the parameter estimates of the VECM model. Table 6 reports the averages of the parameter estimates and the t-values on the error correction coefficients $\gamma \equiv [\gamma_1, \gamma_2]^\top$. The average estimate for γ_1 is negative and that for γ_2 is positive. When the quote from the stock market is high relative to the synthetic quote from the options market, the different signs of the estimates for the two parameters dictate that the stock market quote will adjust downward and the option market quote will adjust upward. Both corrective adjustments move the two sources of quote closer to their long-run equilibrium relation. The average t-values for the two parameters are both large, indicating that both estimates are statistically significant. Quotes from both markets are informative, albeit to different degrees, about the underlying stock price.

Table 6 also reports the parameter averages across the 18 NYSE stocks and across the 22 NASDAQ stocks. The average γ_2 estimates are similar for both NYSE and NASDAQ stocks, but the average γ_1 estimates are much smaller for NASDAQ stocks than for NYSE stocks. This relative difference implies that the options market offers smaller error correcting power on the NASDAQ market quotes than on the NYSE market quotes. This difference matches the difference in the information share estimates, especially on the lower bounds of the information share. For the 18 NYSE stocks, the average lower bound of the information share attributable to the options market is 12.7%, but the average among the 22 NASDAQ stocks is only 3.1%.

[4] During our 2002 sample period, XON and MOB have been merged into one company, XOM.

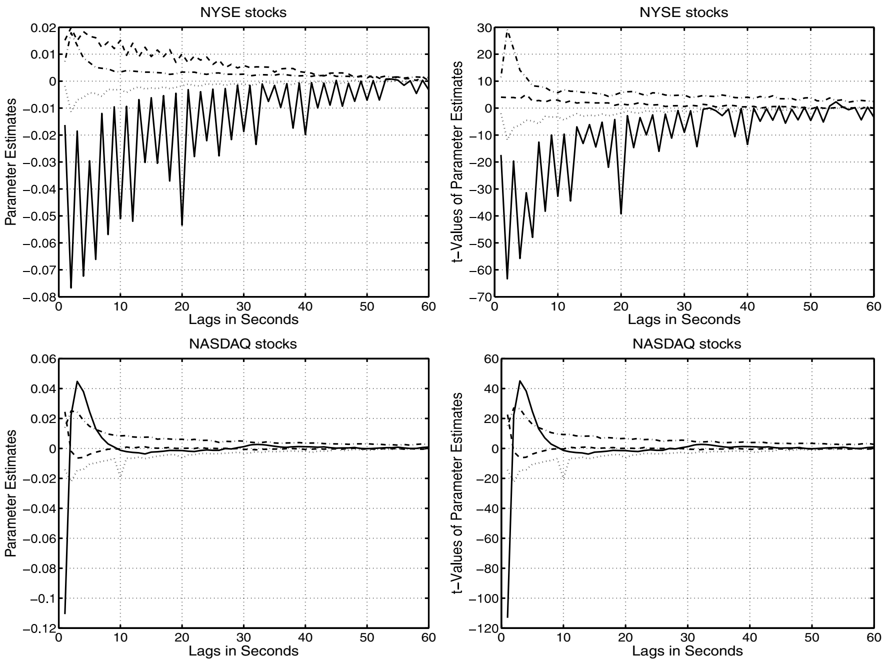

Fig. 1 VECM Parameter Estimates. Lines denote the average estimates for the VECM parameters A_i at different lags $i = 1, 2, \dots, 60$ (seconds). The left panels are the average parameter estimates. The right panels are the averages of the t-values of the parameter estimates. The top panels are averages across NYSE stocks, and the bottom panels are averages across NASDAQ stocks. The solid lines denote the element $A(1, 1)$, the dashed lines denote the element $A(1, 2)$, the dash-dotted lines denote the element $A(2, 1)$ and the dotted line denote the element $A(2, 2)$

The autoregressive coefficients $A_i, i = 1, \dots, 60$ capture the impact of the lagged returns on the forecast of future returns. Figure 1 plots the sample averages of the estimates (left panel) and t-values (right panel) of the four elements of the A matrix as a function of the lags (in seconds). The top panels are averages across NYSE stocks. The bottom panels are averages across NASDAQ stocks.

The solid lines denote the $A(1, 1)$ component, which captures the autoregressive feature of the stock market quote. The average estimates for $A(1, 1)$ on NYSE stocks are mostly negative, with the absolute magnitude declining with increasing lags. The line exhibits a strong zigzagged shape, potentially representing noises induced by bid-ask bounces. The t-value plot shows that these autoregressive coefficients are statistically significant up to 30-second lags.

The average estimates for NASDAQ stocks are much smoother (bottom panels), representing a higher efficiency for the NASDAQ market quotes. The estimates start as negative for one second lag, but become positive at longer lags, and then decay to zero as the lag increases. The decay of the autoregressive coefficients for NASDAQ stocks is faster than for NYSE stocks. The average coefficients for NASDAQ stocks become close to zero for lags over ten seconds.

The $A(2, 2)$ component captures the autoregressive feature of the options market quote. The dotted lines in Fig. 1 denote the estimates and t-values for this element at different lags. For both NYSE and NASDAQ stocks, the average coefficient estimates are negative at all lags. The magnitudes are much smaller than that for $A(1, 1)$, but decay slowly with increasing lags. The coefficients become statistically insignificant when the lags are over 30 seconds.

The cross-impacts are captured by the off-diagonal terms of the coefficient matrix A. The element $A(1, 2)$, shown in Fig. 1 in dashed lines, captures the impact of lagged option market quotes on the stock market quotes. For NYSE stocks, the estimates are all positive and decay slowly with increasing lags. The average t-value is 4.05 at the first lag, and remains above two for lags of up to ten seconds. For NASDAQ stocks, the average estimate for this coefficient is positive for the first lag, but become negative for lags between two and six seconds, after which the coefficients become insignificantly different from zero. Compared to other elements of the A matrix, the estimates for this cross-impact element have small t-values, implying that the information content of the lagged option market returns is limited in forecasting stock returns from the stock market.

The impact of the lagged stock market returns on the option market returns is captured by the estimates for the element $A(2, 1)$ and is represented by the dash-dotted lines. For both NYSE and NASDAQ stocks, the t-values of the estimates for $A(2, 1)$ are larger at short lags and remain over two across all 60 lags. Thus, the impact of the stock market on the option market is not only larger, but also more persistent.

The information share estimates, the common factor loading, and the VECM model parameter estimates show that the price discovery about the future stock price movement occurs more in the stock market than in the options market, more so for NASDAQ stocks than for NYSE stocks. Nevertheless, the information content in the options market is statistically significant for all 40 stocks during our sample period.

2.3 Conditional VECM based on stock price moves

One possible explanation for the low information content in the options market is that transaction costs on options are much higher than transaction costs on stocks. When transaction costs overweigh the benefits of trading with leverage (Black, 1975) and hiding among multiple contracts (Easley, O'Hara, and Srinivas, 1998), the stock market remains the more efficient market for betting on the directional movements in the stock price.

Nevertheless, the benefits will increase and can overweigh the transaction cost for investors with private information that points to large price movements. Thus, we hypothesize that the informativeness of options quotes increases in the presence of large stock price movements. To test this hypothesis, we propose a conditional vector error correction model, with the conditional criterion based on the magnitude of the stock price moves. For each stock, based on the quotes from the stock market, we define log returns over 30-second intervals and compute the tenth and 90th percentiles of these log returns over the 52 business days. To correct for the intraday pattern of return volatilities, the percentiles are computed within each half-hour interval from

9:30am to 16:00. We estimate the following conditional VECM model,

$$\Delta y_{t+1,t+30} = \mathcal{M} + \mathcal{M}[A_1 \Delta y_{t,t-10} + A_2 \Delta y_{t-10,t-20} + A_3 \Delta y_{t-20,t-30} + \gamma(s_{t-1} - o_{t-1})] + u_t, \quad (11)$$

where $\Delta y_{t_i,t_j}$ denotes the log return vector from the two markets between time t_i and t_j, $\mathcal{M}$ is a conditional class variable, which is set to zero when $\Delta y_{t+1,t+30}$ is larger than the tenth percentile but smaller than the 90th percentile, and is set to one when $\Delta y_{t+1,t+30}$ is beyond the tenth or 90th percentiles and hence represents large moves of either direction.

Compared to the VECM specification in (6), Eq. (11) introduces several variations. First, we introduce a conditioning class variable to test how the information flow varies with the market condition. Second, we increase the return horizon from one to 30 seconds for the dependent variable. Using 30-second returns instead of one-second returns increases the stability of the estimation, and reduces the bid-ask bounce effect observed in the previous section. Third, instead of including second-by-second lagged returns as explanatory variables, we aggregate them into 10-second intervals to reduce the number of coefficients and increase the stability of the estimation. We also limit the lags up to 30 seconds based on the estimation results in the previous section.

Table 7 summarizes the estimates for the error correction coefficients on each stock under this conditional VECM. The last three rows report the average across the 18 NYSE stocks, the 22 NASDAQ stocks, and across all the 40 stocks. For all the 40 stocks, the error correction coefficient for the stock market, γ_1, is much more negative for large price moves ($\mathcal{M} = 1$) than for small price moves ($\mathcal{M} = 0$). The difference in magnitude ranges from 1.6 times to more than 26 times. The t-values of the estimates are also larger under large price moves for all but one stock. We have also performed F-tests on the significance of the class variable $\mathcal{M}$ on the γ_1 coefficient. The p-values of the F-tests are all less than 0.00001, indicating strong significance for the class variable. The strong significance of the class variable and the observed difference in the conditional estimates of γ_1 indicate that the stock market quotes react much more strongly to its deviation from the option market quotes when the market experiences large moves ($\mathcal{M} = 1$) than when the market experiences small moves ($\mathcal{M} = 0$).

Compared to the conditional estimates on γ_1, the estimates on γ_2 are larger on average, but its dependence on the large or small price moves is much smaller. For the majority of the stocks (37 out of 40), the γ_2 estimates are larger under large price moves than under small moves; but the difference between the t-values is more ambiguous and varies from stock to stock. Therefore, the option market's correction for its deviation from the equilibrium is much stronger, but less dependent on the market conditions. Nevertheless, when we perform F-tests on the significance of the class variable $\mathcal{M}$ on the coefficient γ_2, the p-values are less than 0.0001 for all stocks, again indicating strong significance even for this weaker dependence on market conditions.

Putting the conditional estimates on the two error-correction coefficients together, we conclude that, under large price moves, the options market's adjustment to the stock market changes little but the stock market's adjustment to the options market

Table 7 Error correction coefficients of the conditional VECM model

	γ_1				γ_2				w_2	
	$\mathcal{M}=0$		$\mathcal{M}=1$		$\mathcal{M}=0$		$\mathcal{M}=1$		$\mathcal{M}=0$	$\mathcal{M}=1$
TICKER	Est.	t-value	Est.	t-value	Est.	t-value	Est.	t-value	Est.	Est.
QQQ	−0.007	−20.4	−0.028	−44.7	0.038	140.7	0.053	107.9	0.156	0.346
TYC	−0.003	−2.8	−0.022	−14.5	0.057	187.9	0.082	199.3	0.052	0.211
GE	−0.002	−4.7	−0.026	−47.3	0.036	161.5	0.053	146.1	0.042	0.329
AOL	−0.003	−8.7	−0.041	−79.8	0.026	117.9	0.052	169.1	0.110	0.440
IBM	−0.004	−8.5	−0.025	−36.8	0.052	183.3	0.068	148.2	0.064	0.268
JPM	−0.008	−4.6	−0.113	−38.4	0.056	176.7	0.069	131.8	0.129	0.623
EMC	−0.007	−17.1	−0.071	−108.2	0.034	128.0	0.059	136.5	0.168	0.545
WMT	−0.005	−12.1	−0.036	−60.2	0.036	133.7	0.060	142.1	0.113	0.374
HD	−0.005	−10.3	−0.134	−197.1	0.040	145.6	0.059	155.8	0.113	0.696
C	−0.007	−17.0	−0.060	−85.8	0.044	156.8	0.073	164.1	0.145	0.450
VZ	−0.004	−9.7	−0.035	−56.3	0.034	118.5	0.084	181.5	0.098	0.292
NEM	−0.001	−3.3	−0.023	−42.4	0.028	135.2	0.047	130.8	0.035	0.328
CPN	−0.003	−8.3	−0.024	−46.5	0.025	101.0	0.045	113.8	0.096	0.347
MO	−0.002	−5.5	−0.038	−83.4	0.027	108.2	0.039	129.3	0.072	0.494
JNJ	−0.005	−10.9	−0.015	−25.9	0.044	139.6	0.065	145.2	0.094	0.192
XOM	−0.006	−15.9	−0.062	−92.7	0.029	117.7	0.058	140.3	0.179	0.517
HAL	−0.002	−5.0	−0.012	−23.1	0.025	99.3	0.072	178.1	0.060	0.139
PFE	−0.004	−10.1	−0.033	−51.8	0.041	163.2	0.061	150.7	0.089	0.351
BA	−0.005	−10.6	−0.060	−95.2	0.033	133.3	0.055	149.3	0.120	0.522
SUNW	−0.008	−25.6	−0.018	−39.8	0.024	94.0	0.050	129.8	0.240	0.263
MSFT	−0.006	−18.4	−0.009	−17.8	0.040	155.0	0.040	98.9	0.132	0.189
CSCO	−0.006	−21.2	−0.012	−25.1	0.030	126.1	0.042	113.2	0.173	0.216
INTC	−0.005	−16.8	−0.011	−21.7	0.031	149.4	0.037	111.0	0.142	0.223
NVDA	−0.006	−16.8	−0.028	−52.2	0.044	141.5	0.040	86.9	0.122	0.414
AMAT	−0.006	−18.8	−0.015	−29.3	0.036	148.1	0.042	111.5	0.144	0.257
ORCL	−0.004	−13.2	−0.015	−29.8	0.032	127.0	0.058	132.0	0.109	0.210
QCOM	−0.005	−14.9	−0.028	−51.8	0.040	145.3	0.050	118.6	0.115	0.357
DELL	−0.005	−17.7	−0.016	−37.9	0.025	125.8	0.034	107.6	0.163	0.327
AMGN	−0.005	−13.0	−0.010	−17.9	0.041	143.3	0.050	112.4	0.101	0.164
JNPR	−0.001	−5.9	−0.006	−15.9	0.015	73.0	0.042	126.6	0.082	0.121
SEBL	−0.006	−14.9	−0.037	−57.6	0.037	122.2	0.054	105.2	0.132	0.406
QLGC	−0.007	−16.7	−0.023	−37.7	0.046	140.3	0.043	85.1	0.127	0.352
VRTS	−0.005	−15.2	−0.029	−52.7	0.034	134.5	0.046	116.5	0.137	0.389
EBAY	−0.004	−10.1	−0.026	−44.3	0.055	147.8	0.056	96.6	0.063	0.313
BRCM	−0.009	−24.6	−0.026	−45.7	0.037	153.8	0.052	135.7	0.193	0.334
XLNX	−0.005	−15.4	−0.018	−33.7	0.034	130.5	0.044	106.5	0.135	0.295
KLAC	−0.006	−17.9	−0.021	−36.6	0.047	163.3	0.042	92.3	0.120	0.331
BRCD	−0.003	−11.7	−0.021	−47.3	0.023	111.2	0.036	101.1	0.114	0.366
NVLS	−0.006	−16.4	−0.023	−43.0	0.037	155.4	0.039	101.7	0.130	0.373
GNSS	−0.002	−6.0	−0.027	−47.0	0.032	106.5	0.059	127.6	0.065	0.313

(*Continued on next page*)

Table 7 (*Continued*)

	γ_1				γ_2				w_2	
	$\mathcal{M}=0$		$\mathcal{M}=1$		$\mathcal{M}=0$		$\mathcal{M}=1$		$\mathcal{M}=0$	$\mathcal{M}=1$
TICKER	Est.	t-value	Est.	t-value	Est.	t-value	Est.	t-value	Est.	Est.
NYSE	−0.004	−9.2	−0.046	−65.9	0.037	139.3	0.061	150.7	0.099	0.395
NASD	−0.005	−16.0	−0.020	−37.7	0.035	133.4	0.046	110.2	0.132	0.298
ALL	−0.005	−12.9	−0.032	−50.4	0.036	136.0	0.053	128.4	0.117	0.342

Entries report the conditional error correction coefficients estimates (Est.) $\gamma = [\gamma_1, \gamma_2]^\top$ and the t-statistics of the estimates (t-value). $M = 0$ denotes normal price movements with the ten and ninety percentiles, while $M = 1$ denotes large price moves beyond the ten and ninety percentiles. The model is specified in Eq. (11). The last two columns report the common factor loading on the options market (w_2) for $\mathcal{M} = 1$ and $\mathcal{M} = 0$. The percentiles are computed based on the thirty-second log returns on the stock market quotes.

increases dramatically, implying that the options market becomes significantly more informative under large price movements.

The dependence of the information flow on the magnitude of the price movement is also shown in the conditional common factor loading estimates, reported in the last two columns of Table 7. We compute the common factor loading directly from the error correction coefficients. The loading on the options market increases markedly from $\mathcal{M} = 0$ to $\mathcal{M} = 1$. The average loading under large price moves is 22.5 percentage points higher than the average loading under small price moves. The difference is 29.7 percentage points for NYSE stocks and 16.6 percentage points for NASDAQ stocks.

Under small market moves, price discovery on the directional movement of the stock price mainly occurs in the stock market, possibly due to the smaller transaction cost in that market. Nevertheless, when informed traders expect large price moves, the benefit of high leverage and the availability of multiple contracts attract them to the options market and significantly increase the price discovery in the options market.

3 Price discovery under changing market conditions

Three years have past since the original choice of our sample period. Over these years, the trading activities in the options market have increased dramatically. According to the Options Clearing Corporation, the daily trading volume on exchange-listed stock options averages around 3 million contracts during May to July of 2002. The average has more than doubled to become 6.5 million contracts by March 2006. To gauge the impact of the changing market conditions on the information flow between the stock and the options markets, we choose March of 2006 as a new sample period, which includes 23 business days. Based on data in this sample period, we repeat our analysis on the Nasdaq 100 index tracking stock, the ticker of which has changed from QQQ to QQQQ. As in three years ago, options on QQQQ is the most actively traded among all stock options.

When we estimate the same vector error correction model as in Eq. (6) using the new sample, the cointegration coefficients are estimated to be $\gamma_1 = 0.00215$ with a

t-statistics of -8.77 and $\gamma_2 = 0.0516$ with a t-statistic of 117.58. Again, both the stock and the options markets are informative about future price movements. When we convert these estimates into the common factor loading, we obtain an estimate of $w_2 = 4\%$ for the options market. The information share estimates are between 0.5% and 9.59%. Therefore, although both markets contain information, the informativeness of the options quotes on the stock price movement has declined over the past three years. Also different from the findings reported in Fig. 1, the coefficient estimates on the lagged returns are mostly zero, showing that the information dissemination speed is much higher now than before.

The declining information content in the options market is puzzling. If transaction cost is the main culprit for the low information content as we have conjectured in the previous section, we would expect the information content in the options market to increase over time as the options market becomes more efficient. Although the equity markets have also undergone a period of transformation, the growth in the options market is definitely more explosive during the past few years than that in the stock market. Therefore, transaction costs alone cannot be the full story.

One possible explanation for the declining information content in the options market is related to the recent microstructural transformations in options market making procedures. Despite the rapid and continuing growth in the options market, the options trading activities are much thinner than that in the underlying stock. Yet, whenever the stock price moves in the stock market, the options market makers are forced to update all the quotes on the hundreds of options on the stock, even if there are no options trading activity. This phenomenon leads to dramatically higher quote-to-trade ratios in the options market than in the underlying stock market. For example, among the OPRA messages in March 2006, less than 6 million are trade messages, but over 19 billion are quote messages, leading to a quote-to-trade message ratio of well over 3000. To alleviate this extremely uneven quoting burden relative to the sparse trading activity, increasingly many options market making firms have developed automated quote updating algorithms that allow them to update the options quotes automatically whenever the stock price moves in the stock market. In the absence of options order flows, these automated updating algorithms imply that the information flow becomes almost purely one directional from the stock market to the options market. Only when there are significant options transactions do the options quotes adjust accordingly and spill the order flow information over to the stock market.

To test this conjecture, we estimate several conditional VECM models with the new data based on different conditioning class variables. As in Eq. (11), the conditional VECMs are specified as,

$$\Delta y_{t+1,t+30} = \mathcal{M}_k + \mathcal{M}_k[A_1 \Delta y_{t,t-10} + A_2 \Delta y_{t-10,t-20} + A_3 \Delta y_{t-20,t-30} + \gamma(s_{t-1} - o_{t-1})] + u_t, \quad (12)$$

where we consider four set of conditioning class variables, $k = 1, 2, 3, 4$, based on four different hypotheses. We discuss each hypothesis and the estimation results in the following subsections.

3.1 Large stock price movements

Repeating what we have done with the earlier sample on the transaction cost hypothesis, we define the first class variable based on the relative magnitude of price movements. We set $\mathcal{M}_1$ to one when $\Delta y_{t+1,t+30}$ is either larger than the 99th percentile or smaller than the 1st percentile of the 30-second return from the stock market, and is set to zero otherwise.

Table 8 reports the parameter estimates under this hypothesis in panel (i). The estimates show that under both small and large movements, both markets react to lagged returns computed from the stock market more strongly than to lagged returns

Table 8 Conditional VECMs on QQQQ in March 2006

	s_t		o_t		s_t		o_t	
(i)	$\mathcal{M}_1 = 0$ (small moves)				$\mathcal{M}_1 = 1$ (large moves)			
Intercept ($\times 10^5$)	0.0679	(1.63)	0.0727	(1.49)	−2.0035	(−6.86)	−1.5878	(−4.64)
$s_{t,t-10}$	0.0355	(11.23)	0.3528	(95.13)	0.6484	(38.17)	0.4766	(23.92)
$s_{t-10,t-20}$	0.0204	(6.48)	0.2117	(57.27)	0.4614	(24.80)	0.3766	(17.26)
$s_{t-20,t-30}$	0.0162	(5.29)	0.1290	(35.80)	0.2461	(13.51)	0.2406	(11.26)
$o_{t,t-10}$	−0.0063	(−2.65)	−0.2497	(−89.99)	−0.0646	(−4.30)	−0.0320	(−1.82)
$o_{t-10,t-20}$	−0.0093	(−4.00)	−0.1465	(−53.67)	0.0389	(2.55)	0.0588	(3.29)
$o_{t-20,t-30}$	−0.0095	(−4.38)	−0.0794	(−31.13)	0.0543	(4.04)	0.1028	(6.51)
$(s_{t-1} - o_{t-1})$	−0.0036	(−2.33)	0.3182	(174.49)	−0.1686	(−15.20)	0.9244	(71.06)
w_2, %	1.12				15.43			
(ii)	$\mathcal{M}_2 = 0$ (Small number of option trades)				$\mathcal{M}_2 = 1$ (Large number of option trades)			
Intercept ($\times 10^5$)	0.0422	(1.02)	0.0488	(1.00)	−3.0557	(−5.70)	−3.7526	(−5.96)
$s_{t,t-10}$	0.0637	(20.36)	0.3753	(102.28)	0.1065	(3.03)	0.2347	(5.69)
$s_{t-10,t-20}$	0.0383	(12.22)	0.2259	(61.48)	0.0903	(2.64)	0.1406	(3.51)
$s_{t-20,t-30}$	0.0251	(8.23)	0.1400	(39.09)	0.0129	(0.39)	−0.0772	(−1.98)
$o_{t,t-10}$	−0.0061	(−2.61)	−0.2469	(−89.38)	−0.1783	(−6.10)	−0.4126	(−12.03)
$o_{t-10,t-20}$	−0.0059	(−2.54)	−0.1430	(−52.56)	−0.1929	(−6.85)	−0.2886	(−8.73)
$o_{t-20,t-30}$	−0.0060	(−2.78)	−0.0760	(−29.94)	−0.0971	(−3.88)	−0.0551	(−1.88)
$(s_{t-1} - o_{t-1})$	−0.0061	(−3.95)	0.3306	(81.79)	−0.0670	(−2.90)	0.3630	(3.40)
w_2, %	1.82				15.57			
(iii)	$\mathcal{M}_3 = 0$ (Low sell/buy pressure)				$\mathcal{M}_3 = 1$ (High sell/buy pressure)			
Intercept ($\times 10^5$)	0.0332	(.80)	0.0381	(0.78)	−0.6169	(−1.88)	−0.9788	(−2.54)
$s_{t,t-10}$	0.0635	(20.14)	0.3764	(101.72)	0.0832	(4.17)	0.3033	(12.97)
$s_{t-10,t-20}$	0.0370	(11.69)	0.2237	(60.19)	0.0963	(5.28)	0.3085	(14.40)
$s_{t-20,t-30}$	0.0226	(7.32)	0.1383	(38.15)	0.1054	(5.97)	0.1913	(9.24)
$o_{t,t-10}$	−0.0056	(−2.37)	−0.2439	(−87.62)	−0.0837	(−5.11)	−0.4602	(−23.92)
$o_{t-10,t-20}$	−0.0061	(−2.60)	−0.1413	(−51.54)	−0.0542	(−3.35)	−0.3007	(−15.83)
$o_{t-20,t-30}$	−0.0063	(−2.88)	−0.0750	(−29.36)	−0.0212	(−1.39)	−0.1447	(−8.09)
$(s_{t-1} - o_{t-1})$	−0.0055	(−3.54)	0.3316	(181.34)	−0.0680	(−5.44)	0.2626	(17.91)
w_2, %	1.64				20.58			

(*Continued on next page*)

Table 8 (*Continued*)

(iv)	s_t		o_t		s_t		o_t	
	$\mathcal{M}_4 = 0$ (Low consistent sell/buy pressure)				$\mathcal{M}_4 = 1$ (High consistent sell/buy pressure)			
Intercept ($\times 10^5$)	0.0256	(0.62)	0.0274	(0.56)	−0.6897	(−1.18)	−1.3838	(−2.02)
$s_{t,t-10}$	0.0632	(20.21)	0.3737	(101.89)	0.1052	(2.99)	0.3464	(8.40)
$s_{t-10,t-20}$	0.0374	(11.94)	0.2247	(61.19)	0.1661	(4.61)	0.2636	(6.23)
$s_{t-20,t-30}$	0.0249	(8.14)	0.1396	(38.97)	0.0664	(1.98)	0.0642	(1.64)
$o_{t,t-10}$	−0.0057	(−2.42)	−0.2446	(−88.53)	−0.1811	(−6.72)	−0.6456	(−20.44)
$o_{t-10,t-20}$	−0.0063	(−2.70)	−0.1428	(−52.48)	−0.1193	(−4.15)	−0.3492	(−10.35)
$o_{t-20,t-30}$	−0.0064	(−2.97)	−0.0757	(−29.81)	−0.0369	(−1.39)	−0.1238	(−3.99)
$(s_{t-1} - o_{t-1})$	−0.0057	(−3.64)	0.3317	(181.89)	−0.0952	(−5.99)	0.1854	(9.95)
w_2, %	1.68				33.93			

Entries report the coefficient estimates of the following conditional error correction models:

$$\Delta y_{t+1,t+30} = \mathcal{M}_k + \mathcal{M}_k[A_1 \Delta y_{t,t-10} + A_2 \Delta y_{t-10,t-20} + A_3 \Delta y_{t-20,t-30} + \gamma(s_{t-1} - o_{t-1})] + u_t,$$

where $\Delta y_{t_i,t_j} \equiv [s_{t_i,t_j}, o_{t_i,t_j}]^\top$ denotes the log return vector from the two markets between time t_i and t_j, $\mathcal{M}_k$ is a conditional class variable. Each panel in the table represents a different specification for the conditioning class variable:

(i) $\mathcal{M}_1$ is set to one when $\Delta y_{t+1,t+30}$ is either larger than the 99th percentile or smaller than the 1st percentile, and is set to zero otherwise.
(ii) $\mathcal{M}_2$ is set to one when the total number of trades on the underlying call and put options within the last 30 seconds is greater than the 99th percentile, and is set to zero otherwise.
(iii) $\mathcal{M}_3$ is set to one when the sell pressure from the underlying call and put options within the last 30 seconds is either larger than the 99th percentile or smaller than the 1st percentile, and is set to zero otherwise. The sell pressure is defined as

$$\begin{aligned}\text{sell pressure} &= \text{number of sell trades on the call} - \text{number of buy trades on the call}\\ &\quad + \text{number of buy trades on the put} - \text{number of sell trades on the put.}\end{aligned}$$

A trade is classified as a buy if it is higher than the mid-quote and a sell if it is lower than the mid-quote. Trades that are at the mid are treated as neither buy nor sell.
(iv) $\mathcal{M}_4$ is set to one when the sell pressure from the underlying call and put options within the last 30 seconds is consistent with the deviation between s_{t-1} and o_{t-1} and is set to zero otherwise. The consistency condition is defined as either the sell pressure is greater than the 99th percentile and $s_{t-1} - o_{t-1}$ is greater than its 75th percentile, or the sell pressure is lower than the 1st percentile and $s_{t-1} - o_{t-1}$ is lower than its 25th percentile.

The estimation is based on QQQQ data during March of 2006.

computed from the options market, consistent with our general observation that the stock market is more informative about the stock price movement.

We also observe that under both small and large moves, the coefficient estimates on the cointegration term are significantly negative for the stock market and significantly positive for the options market, indicating that both markets contain information about the future stock price movement.

Importantly, the coefficient estimates on the cointegration term are larger on both markets conditioning on large moves. When we convert the coefficient estimates into the Gonzalo and Granger (1995) common factor loading, as shown in the last row of the panel, the loading on the options market is merely 1.12% conditioning on small moves, but increases to 15.43% conditioning on large moves, reaffirming that the transaction cost hypothesis accounts for at least part of the low information content in the options market.

3.2 Large number of option trades

To test the hypothesis on the interactions between options trading activities and the information flow, we define a second conditioning class variable based on number of trades on the call and the put option pairs that define the option-implied stock price. At each point in time, we first aggregate the number of trades on the two option contracts during the last 30 seconds. Then, we set the conditioning class variable $\mathcal{M}_2$ to one when the aggregate number of trades is greater than its 99th percentile, and set it to zero otherwise.

The estimation results are reported in panel (ii) of Table 8. In the absence of a large number of options trades ($\mathcal{M}_2 = 0$), the common factor loading for the option market is merely 1.82%. In contrast, when there is a large number of options trades during the last 30 seconds, the loading for the options market increases to 15.57%. The different loading estimates under the two conditions confirm our hypothesis that the options market is informative about the stock price movements only when there are significant options trading activities.

3.3 High sell or buy pressure from the options market

Although the class variable $\mathcal{M}_2$ captures the relative intensity of the options trading, it does not tell us the direction of the order flow. Since the call and put options have opposite exposures on the stock price movement, the impact of a buy order on the call option and the impact of a buy order of similar size on the put option are likely to cancel out each other. To capture the effect of the net order flow, we define a third conditioning variable $\mathcal{M}_3$ based on the net sell pressure from the options market. We define the net sell pressure as,

$$\begin{aligned}\text{sell pressure} \;=\; & \text{number of sell trades on the call} \;-\; \text{number of buy trades on the call} \\ & +\text{number of buy trades on the put} \;-\; \text{number of sell trades on the put}\,.\end{aligned} \tag{13}$$

A negative estimate of the sell pressure indicates a buy pressure on the underlying stock. Based on the algorithm proposed by Lee and Ready (1991), we classify a trade as a buy order if the transaction price is higher than the mid-quote and classify it as a sell order if the transaction price is lower than the mid-quote. Since we are interested in the net order flow, we discard transactions that occur at the mid. Based on this classification, at each point in time, we aggregate the number of buy and sell orders on the call and the put options according to Eq. (13) over the last 30 seconds. Then, we set the class variable $\mathcal{M}_3$ to one if the sell pressure is either larger than its 99th percentile or smaller than its 1st percentile, in which case it is a high buy pressure.

Table 8 reports the estimates in panel (iii). In the absence of high net order flow of either direction from the options market, the common factor loading for the options market is estimated to be 1.64%. On the other hand, when there is high sell or buy pressure from the options market, the loading estimate increases to 20.58%, showing that conditioning on the net order flow is more effective than conditioning on the total number of option trades. Therefore, both the direction and the quantity of the option trading activities are informative about the underlying information flow between the two markets.

3.4 High consistent sell or buy pressure from the options market

Finally, we compare the net order flow with the price deviation between the two markets ($s_{t-1} - o_{t-1}$) to determine whether the sell or buy pressure is consistent with the direction of the price deviation. When there is significant sell pressure from the options market, we expect the stock price inferred from the options market to be lower and hence the deviation ($s_{t-1} - o_{t-1}$) to be positive. On the other hand, if there exists high buy pressure from the options market, we expect the deviation to be negative as the option-inferred stock price (o_{t-1}) is pressured to be higher. When the direction of the deviation matches the direction of the net order flow, we call it consistent sell or buy pressure, in which case we expect the options market to be even more informative.

According to this hypothesis, we set the fourth class variable $\mathcal{M}_4$ to one either when the sell pressure is greater than the 99th percentile and the deviation is greater than its 75th percentile, or when the sell pressure is lower than the 1st percentile and the deviation is lower than its 25th percentile. The class variable is set to zero otherwise.

Panel (iv) of Table 8 reports the estimates under the conditioning of consistent buy or sell pressures from the options market. Without consistent sell or buy pressure, the common loading from the options market remain low at 1.68%. With consistent buy or sell pressure, the options market becomes significantly more informative as the common factor loading reaches as high as 33.93%.

Overall, the estimates from the conditioning VECM specifications show that under the current market microstructure framework where options quoting is largely automated, options quotes contain little new information about the underlying stock price unless there are significant options trading activities. The information content in the options quotes increase further when the options trading activity generates net sell or buy pressure on the underlying stock price, even more so when the pressure is consistent with quote deviation from the two markets.

4 Conclusion

In this paper, we propose a portfolio approach to circumvent the long-standing difficulty in the price discovery study between a primary asset market and its options market. The portfolio combines a long position in a call option with a short position in a put option with the same strike price and maturity. The payoff to this portfolio synthesizes the payoff to a forward contract and is therefore linear in the underlying asset price, even though the payoff function of each individual option is nonlinear. As a result, although the value of an option depends both on the underlying asset price and on return volatilities and higher moments, the value change of this portfolio depends primarily on the change in the underlying asset price level, but not on volatility or other higher moments.

We apply this portfolio approach to the price discovery study between the stock market and the corresponding stock options market in the United States using two months worth of tick data in 2002 and one month of tick data in 2006. We find that despite the growing integration among the different options exchanges and the improvement in the options quoting quality, the price discovery on the directional movement of the stock price still occurs more in the stock market than in the options market, especially for NASDAQ stocks. We also find that the informativeness of the options market declines over time as options market makers adopt automated quoting algorithms that respond automatically to price movements in the stock market. The options quotes become informative only when there are significant options trading activities around the quoting period. In particular, the information content in the options market increases when the options trading activity generates net sell or buy pressure on the underlying stock price, even more so when the pressure is consistent with quote deviations between the stock and the options markets.

This article focuses on the information flow between the stock market and the options market in terms of the directional movement of the underlying stock price. An intriguing line for future research is to apply the portfolio approach to study the information flow and pricing of return variance between the two markets. In principle, we can use a portfolio of plain vanilla options to synthesize a variance swap contract (Carr and Madan, 1998; Carr and Wu, 2004). Then, we can study the dynamic interactions between this synthetic variance swap and the realized variance of the asset returns computed from the stock market. The dynamic relation between the variance swap and the realized variance not only reflects the information flow on the variance of the asset return between the two markets, but also reveals insights on how the market prices the risk of the potentially stochastic return variance.

Acknowledgments We thank Menachem Brenner (the editor), two anonymous referees, Marti Subrahmanyam, Robert Battalio, Joel Hasbrouck, Robert Schwartz, and David Whitcomb for comments. We also thank Automated Trading Desk, LLC for providing the data and computing assistance. We owe the hypothesis on the options trading activities and their effects on the information flow to Menachem Brenner.

References

Andersen, T. G., L. Benzoni, and J. Lund. (2002). "An Empirical Investigation of Continuous-Time Equity Return Models," *Journal of Finance* 57, 1239–1284.

Baillie, R. T., G. G. Booth, Y. Tse, and T. Zabotina. (2002). "Price Discovery and Common Factor Models," *Journal of Financial Markets* 5, 309–321.

Bakshi, G., C. Cao, and Z. Chen. (1997). "Empirical Performance of Alternative Option Pricing Models," *Journal of Finance* 52, 2003–2049.

Bates, D. (1996). "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options," *Review of Financial Studies* 9, 69–107.

Bates, D. (2000). "Post-'87 Crash Fears in the S&P 500 Futures Option Market," *Journal of Econometrics* 94, 181–238.

Battalio, R., B. Hatch, and R. Jennings. (2004). "Toward a National Market System for U.S. Exchange-Listed Equity Options," *Journal of Finance* 59, 933–962.

Bhattacharya, M. (1987). "Price Changes of Related Securities: The Case of Call Options and Stocks," *Journal of Financial and Quantitative Analysis* 22, 1–15.
Black, F. (1975). "Fact and Fantasy in use of Options," *Financial Analysts Journal* 31, 61–72.
Black, F. (1976). "Studies of Stock Price Volatility Changes." In *Proceedings of the 1976 American Statistical Association, Business and Economical Statistics Section* (American Statistical Association, Alexandria, VA.
Black, F. and M. Scholes. (1973). "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy* 81, 637–654.
Booth, G. G., R. W. So, and Y. Tse. (1999). "Price Discovery in the German Equity Derivatives Market," *Journal of Futures Markets* 19, 619–643.
Britten-Jones, M. and A. Neuberger. (2000). "Option Prices, Implied Price Processes, and Stochastic Volatility," *Journal of Finance* 55, 839–866.
Carr, P. and D. Madan. (1998). "Towards a Theory of Volatility Trading." In R. Jarrow, (ed), *Risk Book on Volatility* Risk, New York.
Carr, P. and L. Wu. (2003). "What Type of Process Underlies Options? A Simple Robust Test," *Journal of Finance* 58(6), 2581–2610.
Carr, P. and L. Wu. (2004). "Variance Risk Premia," Working Paper, New York University and Baruch College.
Chakravarty, S., H. Gulen, and S. Mayhew. (2004). "Informed Trading in Stock and Option Markets," *Journal of Finance* 59, 1235–1258.
Chan, K., Y. P. Chung, and W.-M. Fong. (2002). "The Informational Role of Stock and Option Volume," *Review of Financial Studies* 14, 1049–1075.
Chan, K., Y. P. Chung, and H. Johnson. (1993). "Why Option Prices Lag Stock Prices: A Trading-Based Explanation," *Journal of Finance* 48, 1957–1967.
Chu, Q. C., W. L. G. Hsieh, and Y. Tse. (1999). "Price Discovrey in the S&P 500 Index Markets: An Analysis of Spot Index, Index Futures, and SPDRs," *International Review of Financial Analysis* 8, 21–23.
De Jong, F. (2002). "Measures of Contributions to Price Discovery: A Comparison," *Journal of Financial Markets* 5, 323–327.
Dickey, D. A. and W. A. Fuller. (1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association* 74, 427–431.
Ding, Z., R. F. Engle, and C. W. J. Granger. (1993). "A Long Memory Property of Stock Returns and a New Model," *Journal of Empirical Finance* 1, 83–106.
Ding, Z. and C. W. J. Granger. (1996). "Modeling Volatility Persistence of Speculative Returns: A New Approach," *Journal of Econometrics* 73, 185–215.
Easley, D., M. O'Hara, and P. S. Srinivas. (1998). "Option Volume and Stock Prices: Evidence on Where Informed Traders Trade," *Journal of Finance* 53, 431–465.
Eraker, B. (2004). "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices," *Journal of Finance* 59(3), 1269–1300.
Eraker, B., M. Johannes, and N. Polson. (2003). "The Impact of Jumps in Equity Index Volatility and Returns," *Journal of Finance* 58, 1269–1300.
Gonzalo, J. and C. W. J. Granger. (1995). "Estimation of Common Long Memory Components in Cointegrated Systems," *Journal of Business and Economic Statistics* 13, 27–36.
Harris, F. H. deB., T. H. McInish, and R. A. Wood. (2002a). Common Factor Components Versus Information Shares: A Reply," *Journal of Financial Markets* 5, 341–348.
Harris, F. H. deB., T. H. McInish, and R. A. Wood. (2002b). "Security Price Adjustment Across Exchanges: An Investigation of Common Factor Components for Dow Stocks," *Journal of Financial Markets* 5, 277–308.
Hasbrouck, J. (1995). One Security, Many Markets: Determining the Contributions to Price Discovery," *Journal of Finance* 50, 1175–1199.
Hasbrouck, J. (2002). "Stalking the "Efficient Price" in Market Microstructure Specifications: An Overview," *Journal of Financial Markets* 5, 329–339.
Hasbrouck, J. (2003). "Intraday Price Formation in U.S. Equity Index Markets," *Journal of Finance* 58, 2375–2399.
Huang, J. and L. Wu. (2004). "Specification Analysis of Option Pricing Models Based on Time-Changed Lévy Processes," *Journal of Finance* 59(3), 1406–1439.
Jarnecic, E. (1999). "Trading Volume Lead/Lag Relations Between the ASX and ASX Option Market: Implications of Market Microstructure," *Australian Journal of Management* 24, 77–94.

Johansen, S. (1988). "Statistical Analysis of Cointegrated Vectors," *Journal of Economic Dynamics and Control* 12, 231–254.

Johansen, S. (1991). "Estimation and Hypothesis Testing of Cointegrated Vectors in Gaussian Vector Autoregressive Models," *Econometrica* 59, 1551–1580.

Jones, C. S. (2003). "The Dynamics of Stochastic Volatility: Evidence from Underlying and Options Markets," *Journal of Econometrics* 116(1/2), 181–224.

Lee, C. M. C. and M. A. Ready. (1991). "Inferring Trade Direction from Intraday Data," *Journal of Finance* 46, 733–746.

Lehmann, B. N. (2002). "Some Desiderata for the Measurement of Price Discovery Across Markets," *Journal of Financial Markets* 5, 259–276.

Manaster, S. and R. J. Rendleman Jr. (1982). "Option Prices as Predictors of Equilibrium Stock Prices," *Journal of Finance* 37, 1043–1057.

Mayhew, S. (2002). "Competition, Market Structure and Bid-Ask Spreads in Stock Option Markets," *Journal of Finance* 57, 931–958.

Osterward-Lenum, M. (1992). "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics," *Oxford Bulletin of Economics and Statistics* 54, 461–472.

Pan, J. (2002). "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study," *Journal of Financial Economics* 63, 3–50.

Pan, J. and A. M. Poteshman. (2006). "The Information in Option Volume for Future Stock Prices," *Review of Financial Studies* 19, 871–980.

Simaan, Y. E. and L. Wu. (2002). "Price Discovery in the Equity Options Market: An Integrated Analysis of Trades and Quotes," Manuscript, Fordham University.

Stephan, J. A. and R. E. Whaley. (1990). "Intraday Price Change and Trading Volume Relations in the Stock and Stock Option Markets," *Journal of Finance* 45, 191–220.