Price discovery in stock and options markets

Citation for published version:
Patel, V, Putniš, TJ, Michayluk, D & Foley, S 2019, 'Price discovery in stock and options markets', *Journal of Financial Markets*. https://doi.org/10.1016/j.finmar.2019.100524

Digital Object Identifier (DOI):
10.1016/j.finmar.2019.100524

Published In:
Journal of Financial Markets

Price discovery in stock and options markets*

Vinay Patel [a], Tālis J. Putniņš, [a,b] David Michayluk, [a] and Sean Foley [c]

[a] *University of Technology Sydney*

[b] *Stockholm School of Economics in Riga*

[c] *University of Sydney*

Abstract

Using new empirical measures of information leadership, we find that the role of options in price discovery is up to five times larger than previously thought. Approximately one-quarter of new information is reflected in options prices before being transmitted to stock prices, with options playing a more important role in price discovery around information events. Using unique data on traders prosecuted for insider trading, we find that they often choose to trade in options, attracted by their leverage, and when they do the options share of price discovery is higher. Our results help interpret conflicting findings in the existing literature.

Keywords: price discovery, stock, option, information share, insider trading

JEL classification: G14

---

* The Internet Appendix that accompanies this paper is available at goo.gl/FJodvA.
Email: vinay.patel@uts.edu.au, talis.putnins@uts.edu.au, david.michayluk@uts.edu.au, and sean.foley@sydney.edu.au. We thank an anonymous referee, Paul Brockman, John Paul Broussard, Alexander Butler (discussant), Gunther Capelle-Blancard, Carole Comerton-Forde, Robert Daigler, David Easley, Douglas Foster, Lawrence Glosten, Michael Goldstein, Ruslan Goyenko (discussant), Bruce Grundy, Michael Hemler (discussant), Jianfeng Hu, Marc Lipson (discussant), Ron Masulis, Dmitry Muravyev, Maureen O'Hara, Andreas Park, Richard Payne, Tom Smith, Larry Tabb, Terry Walter, John Welborn, Bart Yueshen, and participants at the Society for Financial Studies Cavalcade (Toronto), Financial Management Conference (Paris), Behavioral Finance and Capital Markets Conference (Adelaide), Financial Management Association Conference (Nashville), AFFI EUROFIDAI International Finance Meeting (Paris), Royal Economic Society Conference (Manchester), Financial Management Association Europe Conference (Venice), Australasian Finance and Banking Conference (Sydney), and seminar participants at the University of Rhode Island, University of Tasmania, University of Melbourne, and University of Technology Sydney for helpful comments. We thank the Securities Industry Research Centre of Asia-Pacific for providing access to the data used in this study. Putniņš gratefully acknowledges funding from the Australian Research Council (ARC DE150101889).

1. Introduction

The role of options in price discovery is at the center of an unresolved debate. Theory shows that under reasonable conditions, informed traders will choose to trade in both options and stocks (Back, 1993; Easley et al., 1998), which suggests that options can play a role in price discovery. The importance of this role in price discovery is an empirical question because it is a function of quantities that need to be measured, such as the amount of leverage and liquidity in options.[1] Existing empirical evidence is mixed, and even conflicting. One branch of empirical findings suggests that options might be important in price discovery; for example, abnormal options trading volume and order imbalance can predict future stock returns.[2] In contrast, studies that quantify where price discovery occurs generally attribute a small or even negligible role to actively-traded options, finding information shares as low as 6% in recent times, which is at odds with the ability of options to predict future stock returns.[3] Adding to the confusion, several studies find pervasive abnormal (suspicious) trading activity in options around information events such as takeovers, earnings announcements, and corporate actions.[4] The evidence that news events are frequently preceded by abnormal trading activity in options (suggestive of informed trading) further indicates that information might be impounded through options transactions.

[1] Options markets are attractive for informed traders for many reasons including leverage, advanced trading strategies (e.g., straddles, strips, and spreads), and the ability to circumvent short-sale constraints (Black, 1975; Chakravarty et al., 2004; Lin and Lu, 2016; Hu, 2017).
[2] See Manaster and Rendleman (1982), Anthony (1988), Easley et al. (1998), Pan and Poteshman (2006), Bali and Hovakimian (2009), Johnson and So (2012), Hu (2014), Lin and Lu (2015), and Ge et al. (2016).
[3] See Chakravarty et al. (2004), Holowczak et al. (2006), Czerwonko et al. (2012), Muravyev et al. (2013), and Rourke (2013).
[4] See Amin and Lee (1997), Cao et al. (2005), Acharya and Johnson (2010), Podolski et al. (2013), Augustin et al. (2014, 2015), Chan et al. (2015), Borochin and Golec (2016), Hao (2016), Berkman et al. (2017), Gharghori et al. (2017), and Kacperczyk and Pagnotta (2019).

The lack of consensus raises important questions. First, are options important in price discovery? Is the contribution of options to price discovery different around information events? How often do informed traders choose to trade in the options market and is their information reflected in prices? What drives informed traders' choice of market? In this paper, we address each of these questions. The new evidence in this paper stems from two novel features of our analysis: (1) disaggregating two distinct components of price discovery, which have previously only been considered jointly, and (2) utilizing an extensive and unique dataset of traders prosecuted for insider trading.

Building on the work of Yan and Zivot (2010), we disentangle two components of price discovery in the options market: (1) the relative speed at which options and stock prices reflect new information about fundamental value (termed "information leadership"), and (2) the relative amount of noise in options and stock prices. These two components are obtained by decomposing price movements into permanent and temporary changes. The permanent changes reflect innovations in the fundamental value, whereas the temporary changes occur due to microstructure frictions (e.g., tick size discreteness) or illiquidity (e.g., temporary price pressure that is reversed) and are collectively referred to as "noise." The first component of price discovery indicates where information enters the market and is first reflected in prices (e.g., due to informed trading). This component is the focus of this paper.

In contrast, earlier price discovery measures that have been applied to options markets such as Hasbrouck's (1995) information share (*IS*) and Gonzalo-Granger's (1995) component share (*CS*) capture a mix of both information leadership and the relative level of noise (Yan and Zivot, 2010; Putniņš, 2013). In comparing options and stocks, the magnitude of noise differences between the markets is so large (e.g., options bid-ask spreads are approximately five times larger than those of stocks) that the noise component becomes the primary driver of both *IS* and *CS*, obfuscating the true information leadership component in these metrics. Our central finding is that this obfuscation of the information leadership component due to a high level of noise in options prices is why options appear to make little contribution to price discovery despite evidence that a significant amount of informed trading occurs in options. Using new empirical techniques that isolate the information leadership component [the information leadership share (*ILS*) from Yan and Zivot (2010) and Putniņš (2013)], we show that options do in fact make a meaningful contribution to price discovery.

As a starting point, we use Monte Carlo simulations and empirical proxies for noise to analyze what the various price discovery metrics measure in our setting. We find that conventional measures (*IS* and *CS*) underestimate the informational role of options as a result of the substantially higher level of noise in options prices. This noise is caused by larger tick sizes, less liquidity (resulting in a greater tendency for temporary price pressure), and error in converting options prices to implied stock prices. The simulations show that *ILS*, and an extension we develop in this paper termed the information leadership indicator (*ILI*), are able to capture which market is the relative leader in reflecting new information. In the cases we consider, *ILI* is an approximately unbiased measure of how often a given market leads price discovery, and *ILS* measures the magnitude of the speed differential between the markets.

Using a sample of 35 U.S. stocks with actively-traded options over a ten-year time period, we find that approximately one-quarter of the time new information is impounded into options prices first and then transmitted to stock prices (i.e., average options *ILI* of approximately 29%). This estimate is approximately two to five times larger than previously estimated using *IS* and indicates that an economically meaningful fraction of informed trading occurs in the options market. The relatively low values of *IS* for the options market indicates that options prices are noisier than stock prices, *not* that the options market is redundant in reflecting new information.

We draw on a unique dataset of illegal insider trading prosecutions to supplement our price discovery results. We hand collect every case of insider trading prosecuted by the U.S. Securities and Exchange Commission (SEC) during a 16-year period. We find that 29% of cases and 32% of illegal insider trades take place in options. When insiders trade in both the stock and options markets, the average volume traded in the options market is approximately 50% of their total traded volume. Insider trades in options earn considerably higher percentage profits. For example, insiders that trade only in stocks earn an average of 24%, insiders that trade in both stocks and options earn 39%, and insiders that only trade in options earn an average return of 353%. The substantial number of insider trading cases that involve trading in options supports the notion that informed traders use options markets in addition to stock markets.

Insider trading prosecutions provide a sample of informed trading, which we use to further test whether *ILI* and *ILS* capture information leadership in the two markets. We find that both the options market *ILI* and *ILS*, but neither *IS* nor *CS*, are significantly higher when insiders trade in the options market, supporting the notion that both *ILI* and *ILS* identify where information first enters the market and "detect" the presence of informed trade.

We also find that the share of price discovery in options increases around information events. On days when price-sensitive news is released, the options *ILI* and *ILS* are higher than on non-news days, indicating an increase in the frequency with which options lead stocks and the relative speed of the options market in reflecting new information. This evidence supports the growing number of studies that find abnormal price and/or volume movements in options around a variety of important announcements. Our evidence supports the conjecture that these abnormal patterns are associated with informed trading in the options market.

Next, we investigate *why* informed traders choose to trade in options despite their relative lack of liquidity, and how their trading activity affects the options market. Theory suggests that leverage is one of the main features that attracts informed traders to options (e.g., Easley et al., 1998) and our sample of insider trading prosecutions shows that without margin trading, insider trades in options earn much larger percentage returns than trades in stocks. We analyze how leverage and other factors affect informed traders' choice of market using the options share of price discovery as a measure of the relative amount of informed trading in options. Our analysis faces an endogeneity issue with respect to liquidity; informed traders might choose in which market to trade on the basis of their relative liquidity and this choice is also likely to impact liquidity. To disentangle these effects, we exploit the exogenous reduction in options tick sizes (which occurred in a staggered manner over four years) as an instrument for options liquidity. We find that liquidity has an insignificant bearing on informed traders' choice of market; however, high levels of informed trading in options decrease options liquidity, consistent with increased adverse selection risks. In contrast, we find that leverage draws informed traders to the options market and thus options with higher leverage tend to make a larger contribution to price discovery.

Finally, we use an additional approach to analyzing price discovery, which is based on how prices adjust around episodes in which the prices in the options and stock markets disagree by a sufficient margin so as to imply the presence of arbitrage opportunities. Although this approach is quite different from the vector error correction model (VECM)-based price discovery shares, the results from both approaches indicate that while stocks more often lead options in reflecting new information, options markets do play a role in the price discovery process. Furthermore, the results of both approaches are positively correlated, suggesting they agree as to when options make a relatively larger or smaller contribution to price discovery.

This paper builds on a long line of research that examines the role of options in price discovery. The first branch, theoretical studies, model how and why traders might use options (e.g., Easley et al., 1998). The second branch examines lead-lag relations between options and stocks in terms of returns and volumes, order flow, and implied volatility (e.g., Ge et al., 2016). The third branch uses price discovery measures such as *IS* and *CS*, producing options price discovery estimates as low as 6% for stocks with active options in more recent sample periods (e.g., Muravyev et al., 2013). The fourth branch examines abnormal patterns in options returns, volumes, and implied-volatility around corporate announcements (e.g., Berkman et al., 2017).

2. Data and characteristics of the sample

We employ two samples. The first is 35 large U.S. stocks listed on the NYSE and NASDAQ selected on the basis of having the highest options trading volume in March 2003. This sample spans from April 17, 2003 to April 17, 2013. The second sample, described in more detail in Subsection 4.3, is obtained from prosecuted cases of illegal insider trading in the U.S. It comprises 36 stocks and spans from January 1, 1999 to August 30, 2014.

We obtain intraday trade and quote data for both stocks and options from the Thomson Reuters Tick History (TRTH) database provided by the Securities Industry Research Centre of Asia-Pacific (SIRCA). In our calculation of price discovery measures, we use the National Best

Bid and Offer (NBBO) consolidated quotes for both the stock and options markets.[5] We measure price discovery from midquote prices to minimize the effects of bid-ask bounce. This approach is consistent with the price discovery literature and empirical evidence indicating that the bulk of price discovery occurs through quotes not trades (e.g., Hasbrouck, 1991; Brogaard et al., 2019).

We use intraday data to estimate price discovery measures for each stock-day. Thereafter, we use stock-days as the units of observation. We apply the following criteria to select valid put-call pairs for each stock-day:

(a) Time to maturity (in days) is between one and 70 calendar days (inclusive),

(b) Present value of dividends with ex-dividend dates during the remaining life of the option is less than \$0.05, and

(c) Bid price of the option is greater than or equal to \$0.15.[6]

If there are more than three valid put-call pairs on a given stock-day, we retain only the three pairs with the highest quoting activity.[7]

Table 1 reports descriptive statistics about the activity, liquidity, and leverage of the options and stocks in our sample. The mean time-weighted quoted bid-ask spread (in dollars) for valid put-call options pairs is almost seven times larger than stock spreads, despite lower options prices. Options trading and quoting activity is also lower than that of stocks. Summing across our

---

[5] We do not calculate price discovery metrics when the market is closed in which case midquote prices, spreads, and volume are not meaningful. Dividend information is obtained from CRSP and the continuously compounded risk-free rate is obtained from OptionMetrics.

[6] The purpose of criterion (b) is to reduce the early exercise premium for American call options. This criterion does not remove many stock-days because during the sample period dividends are infrequent and/or mostly less than \$0.05.

[7] The reason for limiting the analysis to a maximum of three put-call pairs per stock-day is that the incremental contribution to price discovery of options beyond that of the three most active pairs is likely to be negligible, and joint estimation of the price discovery measures across multiple put-call pairs becomes exponentially computationally difficult with the number of pairs (due to the need to consider all possible orderings of the prices). In selecting the three put-call pairs each stock-day, when possible, we select one near-the-money option pair satisfying $\left|\ln\left(\frac{S}{K}\right)\right| \leq 6\%$ (where S is the stock price and K is the option strike price) and two away-from-the-money pairs (one with $S > K$, the other with $S < K$) satisfying $6\% \leq \left|\ln\left(\frac{S}{K}\right)\right| \leq 18\%$.

sample of options on each stock-day (up to three put-call pairs per stock-day), the average omega-adjusted dollar volume in options per stock-day is approximately 2% of the stock dollar volume ($10,397,000/$570,401,000).[8] Because our sample includes only a subset of the most liquid/active options contracts each stock-day, the options volume that we report is a lower bound. Despite this, it is clear that there is considerably less volume in options than stocks. The average number of options quote changes per stock-day (summing across put-call pairs) is around one-quarter (25.49%) of the corresponding number in stocks.

< Table 1 here >

The combination of wider bid-ask spreads and lower trading activity in options indicates that options are considerably less liquid than stocks. Therefore options prices are likely to be noisier, impacting the second component of price discovery. Options leverage is measured by omega: the absolute options delta multiplied by the ratio of the stock price to the option price. The leverage in options is clearly evident from the descriptive statistics; average stock prices are approximately 24 times larger than average options prices and the average option omega is 15.90.

[8] To compare options volume with that of the stock market, we have to convert options volume into the same "units" as stock volume, recognizing that because of the leverage in options, one dollar of options gives a much greater risk exposure than one dollar of stocks. Suppose there is an option with a delta of one trading at a price of $2 when the stock trades at $100. Every $2 of options volume has the equivalent dollar exposure as $100 of stock volume (when stocks increase by $1, the options position changes by $1 and so does the stock position). Therefore, options volume would have to be scaled up by $100/$2 = 50 times to make it comparable to stock volume. If instead (more realistically) the option delta is say 0.5, then every $2 of options volume has half (delta times) the dollar exposure of $100 of stock volume (when stocks increase by $1, the options position changes by $1 × 0.5), so options volume would have to be scaled up by $100 × 0.5/$2 = 25 times. This example illustrates that to convert options dollar volume to the stock equivalent in terms of dollar exposure, the options volume has to be multiplied by delta times the ratio of the stock price to the option price, that is, the option omega.

3. Methodology

3.1. Options-implied stock price

Following Muravyev et al. (2013), we begin with the European put-call parity relation to calculate the options-implied stock price for a given put-call pair,

$$S_t = C_t(K,T) - P_t(K,T) + PV_t\big(D(t,T)\big) + Ke^{-r(T-t)}, \quad (1)$$

where S_t is the stock price at time t, $C_t(K,T)$ and $P_t(K,T)$ are the call and put option prices with strike price K and expiry date T, $PV_t(D(t,T))$ is the present value of cash dividends at time t, r is the continuously compounded risk-free rate of interest per annum, and $T-t$ is the time to maturity.[9] We adjust equation (1) to incorporate the ability to exercise early because our sample consists of American-style options. Denoting the early exercise premium by $v_t(K,T)$, we have:

$$S_t + v_t(K,T) = C_t(K,T) - P_t(K,T) + PV_t(D(t,T)) + Ke^{-r(T-t)}. \quad (2)$$

We calculate $v_t(K,T)$ by first estimating the error from the put-call parity relation at every quote update:

$$\varepsilon_t = C_t(K,T) - P_t(K,T) + PV_t(D(t,T)) + Ke^{-r(T-t)} - S_t. \quad (3)$$

The early exercise premium is then calculated as the average error term for each stock-day:

$$v_t(K,T) = \frac{\sum_{j=1}^{N} \varepsilon_j}{N}. \quad (4)$$

Recognizing that we can replicate a stock position using options contracts, we can rewrite equation (2) in terms of the options-implied bid price and options-implied ask price at time t:

$$Implied\ Bid_t(K,T) = C_t^{Bid}(K,T) - P_t^{Ask}(K,T) + PV_t\big(D(t,T)\big) + Ke^{-r(T-t)} - v_t(K,T) \quad (5)$$

---

[9] There are a variety of methods for computing the options-implied stock price. Chakravarty et al. (2004) use the binomial model. Holowczak et al. (2006), Muravyev et al. (2013), and Rourke (2013) use put-call parity. The put-call parity approach has the advantages that it is model-free, relying only on the law-of-one-price or absence of arbitrage, it uses observable parameters only, and it incorporates information from both call and put prices.

$$Implied\ Ask_t(K,T) = C_t^{Ask}(K,T) - P_t^{Bid}(K,T) + PV_t(D(t,T)) + Ke^{-r(T-t)} - v_t(K,T). \quad (6)$$

3.2. Price discovery measures

Commonly-used empirical measures of price discovery include Hasbrouck's (1995) information share (*IS*) and Gonzalo and Granger's (1995) component share (*CS*). Both measures are based on decomposing price changes into permanent components (innovations in the fundamental value) and temporary components (noise). The temporary price changes arise from microstructure frictions, such as tick discreteness and illiquidity, which give rise to short term pricing errors that are subsequently reversed. *IS* decomposes the variance of innovations in the common efficient price and attributes a share of the variance to each price series. *CS* is the normalized weight of the price series in the linear combination of prices that forms the common efficient price. In essence, *IS* is a variance-weighted version of *CS*. Both *IS* and *CS* are calculated from the parameter estimates and reduced form errors of a vector error correction model (VECM) or equivalent vector moving average model (VMA). The *IS* and *CS* measures are improvements on prior lead-lag methods, which do not take into account the cointegration of stock and options prices and do not separate temporary and permanent price changes.

Recent research points out what these price discovery metrics measure and how they differ. Yan and Zivot (2010) and Putniņš (2013) separate price discovery into two components: (1) the relative speed with which a price reflects new information (innovations in the efficient price), and (2) the relative noise in a price series. A price series can be superior to another in the first component only, the second component only, both components, or neither component. Empirical price discovery studies typically aim to measure the first component, relative speed or information leadership, because this component identifies where information enters the market and is first reflected in prices. However, as Yan and Zivot (2010) and Putniņš (2013) show, *IS* and *CS* in fact measure a mix of the two components, often placing considerable weight on the second component, relative noise. Somewhat worryingly, a market that is unambiguously the first to reflect new information about the efficient price (and therefore the leader in price discovery) can have a lower *IS* and *CS* than a related market that is a price follower if the first market has noisier prices (e.g., less liquidity or a larger tick size). Differences in noise can mask information leadership when using *IS* and *CS* to measure price discovery.

Yan and Zivot (2010) show that *IS* and *CS* can be combined in a way that isolates leadership in impounding new information and removes the influence of noise. This is because *IS* and *CS* measure both components of price discovery but in different ratios, so the correct combination of these measures can cancel out their dependence on relative noise.[10] This is analogous to combining two stocks with different betas into a portfolio—with the correct weights, market risk can be purged from the portfolio. Using this approach, Putniņš (2013) defines the information leadership share (*ILS*), which seeks to isolate the relative speed with which a price series reflects new information. *ILS* is easy to interpret and is directly comparable to *IS* and *CS* because it takes the range [0,1] indicating the "share" of price discovery attributable to a given price series. When comparing two series, values above 0.5 indicate that the price series leads in price discovery.

Because *IS* and *ILS* by construction lie in the interval [0,1], in finite noisy samples their expected values will be biased inward from the endpoints of the interval. This occurs due to truncation of the sampling error distribution. To correct for the bias away from zero and one, we define a fourth measure of price discovery, the information leadership indicator (*ILI*). *ILI*, unlike

[10] Yan and Zivot (2010) express information leadership as: $IL_1 = \left|\frac{IS_1}{IS_2}\frac{CS_2}{CS_1}\right| = \left|\frac{(IL_1 N_1)}{(IL_2 N_2)}\frac{N_2}{N_1}\right| = \left|\frac{IL_1}{IL_2}\right|$, $IL_2 = \left|\frac{IS_2}{IS_1}\frac{CS_1}{CS_2}\right| = \left|\frac{(IL_2 N_2)}{(IL_1 N_1)}\frac{N_1}{N_2}\right| = \left|\frac{IL_2}{IL_1}\right|$, where *IS* is found to be a function of the relative speed at which a price reflects new information (*IL*) and the relative noise in the price series (*N*), and *CS* is found to be a function of *N* only. The cross-multiplication of *IL* and *N* results in a metric that captures information leadership only.

the other "shares," is a binary measure equal to one if $ILS > 0.5$ and zero otherwise. Being based on *ILS*, *ILI* embeds Yan and Zivot's (2010) approach to isolating the information leadership component of price discovery. However, unlike *ILS* it is not constrained in its ability to take values of zero or one. *ILI* exploits the fact that *ILS* is unbiased around 0.5, that is, when one market is faster than the other, its expected *ILS* is greater than 0.5 irrespective of the level of noise (Putniņš, 2013), which is not the case for *IS* or *CS*. Intuitively, *ILS* identifies the leader correctly on average but understates the contribution of the leader, whereas *ILI* takes the leader identified by *ILS* and assigns full credit for the leadership. Put simply, there are two forms of bias in *IS* and *CS* caused by noise: (1) bias towards zero for the noisier market and (2) bias inwards from the zero/one endpoints. *ILS* corrects for the first form of bias, while *ILI* corrects for both forms of bias.

In simulations below, we show that *ILI* is an approximately unbiased measure of information leadership, even in finite noisy samples. When the binary *ILI* is averaged across multiple stock-days, it becomes a "share," indicating the proportion of stock-days in which options lead stocks. When used in such a way, *ILI* still displays good finite sample properties.

We estimate the price discovery measures for each stock-day. A stock-day can have up to three valid put-call pairs, producing up to three options-implied stock prices in addition to the actual stock price. We estimate the options price discovery measures by jointly considering the price discovery that occurs in all of the valid put-call pairs. Consequently, we use the multiple-market implementations of the *IS* and *CS* metrics and derive the multiple-market extensions of *ILS* and *ILI*. Below we present the approach for four prices (one stock price and three options-implied stock prices), noting that the approach is similar for stock-days with only one, two, or *N* valid put-call pairs.

Each stock-day, we estimate a reduced-form VECM of log stock midquote prices ($p_{1,t}$) and log options-implied stock midquote prices ($p_{2,t}$ to $p_{4,t}$), with 200 lags (prices are sampled at one-second intervals during continuous trading (09:45AM to 3:45PM)):

$$\Delta p_t = \alpha Z_{t-1} + \sum_{i=1}^{200} b_i \Delta p_{t-i} + e_t, \tag{7}$$

where Δp_t is the 4×1 midquote return vector, α is the $4 \times (4-1)$ matrix of error correction coefficients, Z_{t-1} is the 4×1 co-integrating vector, b_i is the 4×4 coefficient matrix for lag i, and e_t is the 4×1 vector of residuals. Sampling quotes in one-second intervals is consistent with the literature and also finds support in our empirical analysis as an appropriate frequency.[11]

We calculate IS_1 to IS_4, and CS_1 to CS_4, from the VECM estimates using an approach equivalent to Hasbrouck (1995) and Gonzalo and Granger (1995) and detailed in Appendix A. Market i's propensity to reflect new information (how much market i's price responds to an innovation in the efficient price) is given by the ratio $\beta_i = IS_i / CS_i$ (Yan and Zivot, 2010). Normalizing the information leadership propensities as per Putniņš (2013) gives *ILS*:

$$ILS_1 = \frac{\beta_1^2}{\beta_1^2+\beta_2^2+\beta_3^2+\beta_4^2},\ ILS_2 = \frac{\beta_2^2}{\beta_1^2+\beta_2^2+\beta_3^2+\beta_4^2},\ ILS_3 = \frac{\beta_3^2}{\beta_1^2+\beta_2^2+\beta_3^2+\beta_4^2},\ ILS_4 = \frac{\beta_4^2}{\beta_1^2+\beta_2^2+\beta_3^2+\beta_4^2}. \tag{8}$$

From *ILS*, we compute *ILI* as follows:

$$ILI_i = \begin{cases} 1 & \text{if} \quad ILS_i > ILS_k \quad \forall\ k \neq i \\ 0 & \text{otherwise} \end{cases}. \tag{9}$$

[11] Using a one-second sampling frequency and 200 lags in the VECM is similar to previous studies (Chakravarty et al., 2004; Czerwonko et al., 2012; Muravyev et al., 2013). In support of using a one-second sampling frequency, we subsequently show using cross-correlations of stock and options midquote returns that options lead stocks in reflecting new information by up to ten seconds (see Figure 1). The high sampling frequency (one-second) is required to reduce contemporaneous correlation between the price series, which arises due to time aggregation. If contemporaneous correlation between prices is an issue, we will observe a wide spread between the upper and lower bounds of our information share estimates (Baillie et al., 2002). The actual bounds that we estimate are relatively narrow, again suggesting that the one-second resolution of the data is sufficiently granular for attributing information shares.

The options price discovery shares are the sum of the price discovery shares of the options-implied stock prices. For example, the options *IS* is given by the sum of IS_2 to IS_4, while the stock's price discovery share is IS_1 (equal to one minus the options price discovery share). The options *CS*, *ILS*, and *ILI* are obtained in a similar manner.

Throughout the paper we report and analyze the price discovery measures for the options market (the stock market measures are just one minus the option market measure). We scale up all of the price discovery measures to the range 0–100%. The *IS* and *CS* results are qualitatively similar throughout the paper and Internet Appendix. Given that *IS* is more frequently used than *CS*, for brevity we focus our analysis on the differences between *ILS*/*ILI* and *IS*.

4. Results

4.1. What components of price discovery do the empirical metrics capture?

Before using the price discovery measures to quantify the role of options in price discovery, we analyze to what extent they measure information leadership versus relative noise in the specific setting of stock and options markets. We start with Monte Carlo simulations using several different structural models calibrated to our empirical setting, and then turn to the empirical data. In all specifications, we find that *ILI* has good finite sample properties and provides an (approximately) unbiased measure of which market leads in price discovery using samples of similar size and noise as the empirical data. As a continuous measure, *ILS* picks up on the *magnitude* of the speed differential between the two markets. We find that the difference between *IS* and both *ILI*/*ILS* is due to the sensitivity of *IS* to the relative noise (illiquidity) between the two markets, consistent with Yan and Zivot (2010) and Putniņš (2013).

The details of the baseline Monte Carlo simulations are in Appendix B.1 and simulations of a number of different structural models are reported in the Internet Appendix (Section 2). Here we provide a summary of the baseline simulations and results.

We simulate the fundamental value of the stock as a geometric Brownian motion with parameters matching those of a typical U.S. stock. We assume the actual stock price tracks the fundamental value with a delay of δ_1 periods and with a fixed level of noise (pricing errors caused by illiquidity, price discreteness, and so on). We further assume that call and put prices follow the Black-Scholes model, tracking the fundamental value with some other delay δ_2 and noise drawn from $N\left(0, \sigma_{s_2}\right)$. By construction, some days (a fraction q) options lead stocks ($\delta_1 > \delta_2$) and other days (fraction $(1 - q)$) stocks lead options ($\delta_1 < \delta_2$). We vary the parameter q across simulations so that options lead stocks in price discovery from none of the time ($q = 0$) through to all of the time ($q = 1$). We also vary the noise in options prices (σ_{s_2}) across simulations (holding the noise in stock prices fixed) to assess the impact of the relative amount of noise in the two markets. For each set of parameter values, we simulate 5,000 samples of data (each sample corresponding to a day), estimate the options price discovery measures for each sample, and compute the averages across the 5,000 samples.

Table 2 reports the results of the simulations. Across rows we vary the amount of time that options lead stocks in price discovery, from $q = 0$ (none of the time) to $q = 1$ (all of the time). Across columns we vary the amount of noise in options prices relative to stock prices. Bold numbers indicate options price discovery estimates that are greater than 50%.

If the price discovery metrics measure which market impounds new information first irrespective of differences in relative noise levels, we would see no bold numbers in the top half of each 11×6 grid (where $q < 0.5$, meaning options usually follow stocks in reflecting new information) and bold numbers all throughout the bottom half of the grid (where $q > 0.5$, meaning options usually lead stocks). Table 2 shows that only options *ILS* and *ILI* (Panels B and C) satisfy this property and have expected values greater than (less than) 50% when options lead (do not lead) stocks. In contrast, Panel A shows that *IS* takes low values overall, failing to recognize the informational role of options even when by construction (in the simulated data) options are usually the first to reflect new information (bottom half of grids). The low values of *IS* result from its downward bias for the noisier price (in this case options) consistent with Yan and Zivot (2010) and Putniņš (2013). In settings where both markets have similar levels of noise, the bias in *IS* might be small or even negligible, but that is certainly not the case when comparing options and stocks.[12]

Further, if a price discovery measure provides an unbiased estimate of the proportion of time (e.g., fraction of days) that options lead stocks in price discovery, its expected value will be equal to the parameter q, which we vary across rows. Table 2 shows that only *ILI* satisfies this property and provides an (approximately) unbiased measure of price discovery in our setting, irrespective of the level of noise. We therefore focus on *ILI* when quantifying how often options lead stocks in price discovery.

A useful feature of *ILS* is that (unlike *ILI*) it provides information about the *magnitude* of the speed differential between two markets. In additional simulations (reported in Table IA.2b of the Internet Appendix), rather than varying the proportion of time that options lead stocks (q), we vary the magnitude of the speed differential between the markets, $\delta_1 - \delta_2$. The simulations show

[12] Table 2 shows that *IS* is influenced by both information leadership *and* relative noise—holding relative noise fixed, options *IS* decreases when options lead stocks less often (moving vertically up through the grid), and holding information leadership fixed, *IS* decreases when there is more noise in options prices (moving horizontally across the grid from left to right). We confirm that the low values of *IS* (all less than 50%) are due to the high level of noise in options prices by repeating the simulations with considerably lower levels of noise in options prices [using the same levels of noise as stock prices, i.e., $\sigma_{s_2} \in \{0,1,\dots,10\}$ basis points] and find that options *IS* is much higher overall, with approximately half the values above 50%.

that options *ILS* takes larger values when options lead by a larger margin, whereas *ILI* has an expected value close to 100% whenever options lead by even the narrowest of margins. In fact, the *ILS* of a market that leads price discovery approaches 100% as the speed differential (number of periods by which the market leads) becomes large. We therefore report both *ILS* and *ILI* throughout the paper (and *IS* for comparison), noting they measure different quantities—*ILI* being an approximately unbiased measure of the proportion of time that options lead stocks, whereas *ILS* measures the magnitude of the speed differential.

We show that our findings are not unique to the structural models described in this section; a range of different simulated structural models support the conclusions above about the components of price discovery captured by *IS*, *ILS,* and *ILI.* The additional simulation results are reported in Table IA.2a of the Internet Appendix and include the structural models used in Hasbrouck (2002) and Grammig and Peter (2013). The various alternative structural models differ to those described above in the following ways: (i) price discovery takes place through trading on private information, (ii) the fundamental value has a public information component as well as a private information component, (iii) the fundamental value is driven by the level of market volatility, (iv) the noise of each price series is serially correlated, and (v) speed differentials between markets are varied.

< Table 2 here >

Next, we test whether noise impacts the price discovery metrics in the empirical data in a similar manner to what it does in the simulations. According to the analytical results of Yan and Zivot (2010) and our simulations, *IS* underestimates options price discovery due to the higher level of noise in options prices. If this is true, we should see the difference between ILI and IS increase with the relative magnitude of noise in options prices. To test this, we estimate several proxies for the level of noise in the stock and options prices. The first set of noise proxies are calculated as the average absolute deviation of prices from the estimated common efficient price (see Appendix B.2 for details). We sort stock-days into deciles of the differential $(ILI_{it} - IS_{it})$, and for each decile we compute the mean $OptionNoiseRatio_{it}$ (the ratio of options price noise to the sum of options and stock price noise).

Panel A in Table 3 reports the results. As expected, the difference between ILI and IS is positively related to the relative amount of noise in options prices; $OptionNoiseRatio_{it}$ increases almost monotonically from 0.52 in the first decile of $(ILI_{it} - IS_{it})$ to 1.00 in the ninth and tenth deciles. We observe the opposite relation between $(ILI_{it} - IS_{it})$ and $StockNoiseRatio_{it}$ (the ratio of stock price noise to the sum of options and stock price noise). This evidence is consistent with our simulation results and the analytical results in Yan and Zivot (2010), that higher levels of noise in options prices cause IS to underestimate how often options are the first to impound new information.

Multivariate regressions support the conclusion that the more noise there is in options prices, the larger the downward bias in the options IS, creating a larger wedge between the options market ILI and IS. We regress $(ILI_{it} - IS_{it})$ on our measure of relative noise $(OptionNoiseRatio_{it})$ and several control variables including dummy variables for the removal of the grandfathering provision and the options market maker (OMM) exemption (DV_t^{GF} and DV_t^{OMM}), a time trend ($Trend_t$), and stock fixed effects.[13] Model 1 in Panel B of Table 3, shows

[13] The removal of the grandfathering provision and OMM exemption were a part of Regulation SHO. Regulation SHO was introduced on January 3, 2005 to restrict naked short-selling of stocks unless the broker could locate or make arrangements to borrow the stock and deliver to the buyer by the delivery date. If the seller does not borrow the stock in time to close out their position the seller fails to deliver (see Putniņš, 2010). The grandfathering

there is a statistically significant positive relation between $(ILI_{it} - IS_{it})$ and $OptionNoiseRatio_{it}$. A one percentage point increase in $OptionNoiseRatio_{it}$ is associated with a 1.84 percentage point increase in the spread between *ILI* and *IS*. The results are robust to variants of this noise measure: $OptionNoise_{it}$ and $OptionStdNoise_{it}$ (see Appendix B.2 for definitions).

< Table 3 here >

The results are also robust to other model-free measures of relative noise that do not rely on estimating the common efficient price. For example, in Model 2 we use the bid-ask spread in the options market normalized by the sum of the options and stock spreads ($OptionSpread_{it}$) and in Model 3 we use the options market tick size also normalized by the sum of the options and stock tick sizes ($OptionTickSize_{it}$). Across all measures, more noise in options consistently leads to a larger downward bias in the options *IS* (this downward bias is also present in *CS*). We obtain similar results if we replace the $(ILI_{it} - IS_{it})$ differential with the $(ILS_{it} - IS_{it})$ differential (reported in Table IA.3 of the Internet Appendix).

4.2. Price discovery in the stock and options markets

The previous subsection shows that *ILI* is an approximately unbiased measure of which market is the first to reflect new information in a variety of scenarios, while *IS* exhibits a downward bias for the noisier market (options). *ILS* measures the magnitude of the speed

provision for naked short sales was removed on October 15, 2007. The grandfathering provision allowed brokers to naked short-sell the underlying stock, creating fail-to-deliver positions up until the stock got placed on a threshold list. The removal of the options market maker (OMM) exemption occurred on September 17, 2008. This exemption allowed OMMs to hedge options positions by naked short selling the underlying stock.

differential between the markets, approaching 100% only when a market is much faster. In this subsection, we estimate the price discovery measures on our sample to quantify the role of options in price discovery and examine the extent to which conventional measures underestimate price discovery in options.

Panel A in Table 4 reports the means and medians of options price discovery across the 54,714 stock-days in our sample of 35 large stocks with actively-traded options during a ten-year period. Consistent with prior studies, the stock market is the dominant venue for price discovery—the mean *IS*, *ILS*, and *ILI* estimates for options are 11.84%, 36.75%, and 28.94%, respectively.[14] Given what our simulations reveal about the properties of the price discovery measures, the interpretation of *ILI* is the most straightforward: 28.94% of the time new information is impounded into options prices first and then transmitted to stock prices. The bootstrapped 95% confidence interval for the mean *ILI* estimate is (25.90%, 44.58%), indicating a high degree of confidence that options lead stocks on a considerable fraction of stock-days—one-quarter of stock-days at the lower end of the confidence interval and close to one-half at the upper end.

The mean estimates of *IS* are approximately two times smaller than *ILI*, consistent with the fact that noisier options prices cause a downward bias in option *IS*. This result is consistent with both our simulations and our empirical analysis of what explains the difference between *ILI* and *IS*. The mean *ILS* estimate of 36.75% indicates that the average speed differential between the two markets is small, on the order of one to two periods (seconds) in favor of stocks.[15] The speed

[14] We obtain similar results if we estimate price discovery between the stock price and one option price series and calculate the pooled means across put-call pairs: options *IS*, *ILS*, and *ILI* are 5.59%, 41.01%, and 38.20%, respectively.

[15] This conclusion is based on the simulations reported in Table IA.2b of the Internet Appendix in which we vary the speed differential between the markets. An *ILS* of 36.75% falls between the row where options and stocks have equal speed and the row where stocks lead options by two periods.

differential is not constant; *ILI* indicates that one-quarter of the time, the differential is in favor of options and three-quarters of the time it is in favor of stocks (*average ILS* gives the *average* speed differential).

< Table 4 here >

Temporarily digressing from the price discovery metrics, to get a further sense of the speed differentials between the markets, we examine simple lead-lag cross-correlations between the markets on stock-days when options lead ($ILI = 1 = 100\%$) compared to stock-days when stocks lead ($ILI = 0 = 0\%$). Figure 1 shows the cross-correlations (vertical axis), computed as the correlation of one-second midquote returns, r_t^{STOCK} and r_{t+l}^{OPTION}, for each of the lead/lag values (on the horizontal axis) $l = -10, -9, \dots, +9, +10$. For example, the cross-correlation difference at lag $l = -1$ of 0.007 (with associated t-statistic of 10) indicates that the average correlation between r_t^{STOCK} and r_{t-1}^{OPTION} is 0.007 higher when options lead than when stocks lead, and the difference is statistically significant. The cross-correlation difference is larger when we instead compare 1,000 of the $ILI = 1$ stock-days that have the highest *ILS*, and 1,000 of the $ILI = 0$ stock-days that have the lowest *ILS* (see Figure IA.1 of the Internet Appendix). For example, the cross-correlation difference at lag $l = -1$ is 0.031 (with a t-statistic of 12). The distinct pattern in Figure 1 indicates that when $ILI = 1$, options lead stocks by one to ten seconds on average. *ILS* and *ILI* differ from the simple lead-lag illustration in Figure 1 in that these measures focus only on permanent price changes (not temporary ones) and aggregate leadership across all lags.

< Figure 1 here >

Returning to the price discovery metrics in Table 4, our estimate that options prices lead stock prices about one-quarter of the time is between two to five times larger than previously measured using *IS*.[16] Our larger estimates indicate that a meaningful fraction of price discovery occurs in the options market, consistent with theoretical models and a significant empirical literature reporting informed trading in options. Most of the difference compared to previous studies comes from the fact that *ILI* corrects the downward bias in *IS*. However, a second methodological difference is that we jointly consider price discovery across multiple options (up to three put-call pairs per stock-day), whereas prior studies examine price discovery in individual options or single put-call pairs [an exception being Rourke (2013)].

The differences between the upper and lower bounds of *IS* (denoted as *UML*) are fairly small (mean of 1.74%), indicating that the sampling frequency is sufficiently high to reliably attribute price discovery to each price series and avoid excessive contemporaneous correlation. If options and stock prices simultaneously reflect new information (which can arise when the sampling frequency is too low relative to the speed of price adjustment), this would manifest as a wide spread between the upper and lower bound estimates of *IS*.

The pooled means of the price discovery measures are higher than their medians, suggesting the distribution of options price discovery is right-skewed (Table 4, Panel A). This is consistent with the notion that on a small proportion of stock-days (e.g., days on which important news is released), options make a very large contribution to price discovery, while on the majority of

[16] Chakravarty et al. (2004) find that near-the-money options have an *IS* of 17% between 1988 and 1992. Holowczak et al. (2006) and Muravyev et al. (2013) find that the options *IS* is approximately 11% during 2002 and 6.25% between 2003 and 2006, respectively. Rourke (2013) considers options *IS* jointly across multiple strike prices and obtains an estimate of 17% between 2007 and 2008.

stock-days (e.g., non-informational stock-days), options make a considerably smaller contribution to price discovery. In subsequent sections we investigate the role of options around information events and find support for this conjecture.

Panel B in Table 4 reports the standard deviations of the price discovery estimates across the full sample of stock-days (σ_{Total}). The standard deviations are large; for example, $\sigma_{Total} = 40.56\%$ for *ILS* and $\sigma_{Total} = 45.17\%$ for *ILI*. There are two sources of this variation: (1) sampling error in estimating the price discovery measures, and (2) true variation in price discovery across stock-days, which could be variation through time, variation across stocks, or both. To understand what drives the high standard deviations of *ILS* and *ILI* estimates, we decompose the standard deviations into sampling variation (σ_{ε}) and variation in price discovery (σ_{PD}). Assume that a price discovery estimate on a given stock-day is the sum of the true price discovery value and an error, $PD_{it}^{MEASURED} = PD_{it}^{TRUE} + \varepsilon_{it}$. It follows that:

$$Var\left(PD_{it}^{MEASURED}\right) = Var\left(PD_{it}^{TRUE}\right) + Var(\varepsilon_{it}) + 2Cov(PD_{it}^{TRUE}, \varepsilon_{it}). \tag{10}$$

We observe $Var\left(PD_{it}^{MEASURED}\right)$; it is the square of the standard deviation of the price discovery point estimates (σ_{Total}^2). We estimate $Var(\varepsilon_{it})$ from the bootstrap applied to each stock-day (details of the bootstrap are in Appendix A), taking the variance of the distribution of sampling errors. The third term, $2Cov(PD_{it}^{TRUE}, \varepsilon_{it})$, is zero for unbiased measures of price discovery. Therefore, for the approximately unbiased *ILI*, we can rearrange equation (10) to obtain $Var\left(PD_{it}^{TRUE}\right) = Var\left(PD_{it}^{MEASURED}\right) - Var(\varepsilon_{it})$. Performing this decomposition and converting the variance terms to standard deviations, we find that most of the total variation in *ILI* ($\sigma_{Total} = 45.17\%$) is due to true variation in information leadership ($\sigma_{PD} = 44.93\%$), with only a relatively small contribution from estimation/sampling error ($\sigma_{\varepsilon} = 4.67\%$).

In decomposing the variance of *IS*, we know from our simulations that they underestimate options price discovery and the magnitude of the negative error term increases with PD_{it}^{TRUE}, implying $Cov\left(PD_{it}^{TRUE}, \varepsilon_{it}\right) < 0$. Therefore, for *IS*, $Var\left(PD_{it}^{TRUE}\right) > Var\left(PD_{it}^{MEASURED}\right) - Var(\varepsilon_{it})$ and thus $\sqrt{Var(PD_{it}^{MEASURED}) - Var(\varepsilon_{it})}$ (which we label σ_{PD}) gives a lower bound of the true variation in information leadership. Consistent with that notion, we obtain considerably lower estimates of $\sigma_{PD} = 8.81\%$ for *IS*. From the simulations, we can infer that for *ILS*, $Cov\left(PD_{it}^{TRUE}, \varepsilon_{it}\right)$ is also negative but smaller in magnitude than for *IS*. Consistent with this notion, σ_{PD} obtained from *ILS* (39.89%) is closer to the true σ_{PD} obtained from *ILI*.[17]

We examine the extent of variation in price discovery separately across stocks and through time. Panel C in Table 4 reports mean price discovery estimates by ticker. There is a moderate level of cross-sectional variation; the means for different stocks range between 7.94%–18.41% for *IS*, 21.44%–52.18% for *ILS*, and 13.88%–47.12% for *ILI*.

The time series of options price discovery estimates provides insight on the effects of institutional changes in stock and options markets. In Figure 2, we plot the mean *IS*, *ILS*, and *ILI* through time. Options *ILS* and *ILI* exhibit similar trends throughout the sample period. The options *ILI* (solid double line) increases from around 30% in 2003 to 40% in 2007. The increase coincides with increased options trading volume (relative to stock volume) over this period. Figure IA.3 in the Internet Appendix shows that options trading volume (solid black line) grew at a rate of four times the growth in stock trading volume (dotted line) during the period 2003–2007. The rapid growth in options trading volume corresponds with the multi-listing of options

[17] If we instead consider *IS* and *ILS* as measures of different aspects of price discovery rather than biased measures of information leadership (implying the term PD_{it}^{TRUE} is the mix of information leadership and noise in the case of *IS*, and a measure of the speed differential in the case of *ILS*), then $Cov(PD_{it}^{TRUE}, \varepsilon_{it}) = 0$ because *IS* and *ILS* are unbiased measures of themselves. Viewed this way, the results show that there is more variation in the information leadership component of price discovery (i.e., σ_{PD} for *ILI* and *ILS*) than an aggregate of information leadership and relative avoidance of noise (i.e., σ_{PD} for *IS*).

across different exchanges, the introduction of the International Securities Exchange in 2000 and the Boston Option Exchange in 2004, the introduction of electronic option trading platforms, and increased market volatility. Additionally, the removal of the grandfathering provision on October 15, 2007 is likely to have increased the use of options to create synthetic short positions.

< Figure 2 here >

Figure 2 shows that between 2007 and 2013, the options *ILI* declines to approximately 20%. This decline coincides with several institutional changes. The first is the removal of the short sales uptick rule on July 6, 2007, making it easier to execute short sales, potentially reducing the use of options to take bearish positions.[18] The second is the removal of the OMM exemption on September 17, 2008, which increased the cost of short selling for OMMs, likely reducing options market liquidity and price discovery. The third is increased fragmentation in the options market, both across a larger number of venues and number of contracts, as well as via an increased use of autoquotation algorithms. For example, the introduction of the C2 options exchange and weekly contract expiry dates in 2010, and the launch of NASDAQ OMX BX and MIAX option exchanges in 2012. For the stocks in our sample, Figure 2 shows that the mean number of put-call pairs per stock at a given point in time satisfying the criteria in Section 2 (solid grey line) increases from around 5 to 12 between 2007 and 2013.

In summary, the results in this section indicate that options play a more important role in price discovery than previously thought; approximately one-quarter of the time new information

[18] The uptick rule restricts short selling to the following circumstances: (i) on an uptick (at a price greater than the last traded price), or (ii) on a zero-plus tick (at the last traded price if the last trade was made on an uptick). We do not include a dummy variable for the removal of the uptick rule in our regressions because it occurs close to the time of the removal of the Grandfathering provision and therefore has a correlation of 0.92 with DV_t^{GF}.

is reflected in options prices before being transmitted to stock prices. Using *ILI*, this estimate is several times larger than the options *IS* documented in prior studies. Our simulations and empirical analysis of the price discovery metrics show that the relatively high level of noise in options prices masks their role in impounding new information in measures such as *IS*.

4.3. Insider trading prosecutions

Given the evidence that options lead stocks in impounding new information approximately one-quarter of the time, next we examine the drivers of this informational role. We analyze how often informed traders choose to trade in the options market and what effect this choice has on the information reflected in prices. For this we turn to prosecuted cases of insider trading. The insider trading cases provide additional evidence on whether informed traders use options markets, and to test the ability of the price discovery metrics to "detect" the presence of informed trading in options.[19]

We start by examining all 7,061 SEC litigation releases relating to insider trading between January 1, 1999 and August 30, 2014.[20] Within these cases, we identify 539 news announcements that are preceded by illegal trading on insider information (hereafter "insider trading"). The cases involve instances in which the illegal trading occurs in options only, stocks only, and in both markets.

In the majority of cases, insiders trade in the stock market (382 of 539 announcements), a significant proportion of announcements (157 of 539, or 29%) involve insider trading in the

[19] Several studies examine insider trading in the U.S. stock markets and generally find that insider trading leads to more rapid price discovery (e.g., Meulbroek, 1992; Augustin et al., 2014; Ahern, 2017; Kacperczyk and Pagnotta, 2019).

[20] SEC litigation releases are obtained from www.sec.gov/litigation/litreleases.shtml. Our sample contains a variety of different news announcement types including mergers, analyst recommendation changes, and earnings announcements. The majority of announcements (85%) contain positive news about the value of the company.

options market either exclusively or in addition to trading in the stock market. Closer examination of the insider trades and volumes provides further evidence that insiders often take advantage of their private information using the options market. In total, 32% of insider trades are in options (739 of 2,320). Across all announcements, the average share volume that insiders trade in the options market is 22.94% of their total traded volume, or 52.14% when insiders trade in both stocks and options. The SEC prosecutions indicate that a meaningful fraction of insider trading occurs in the options market. In subsequent analysis, we examine whether options prices, and thus the options share of price discovery, reflects trading by insiders.[21]

Consistent with theoretical predictions, the inherent leverage in options magnifies insider traders' percentage profits. Insider trades in the stock market only earn an average of 24%, whereas those that trade in both the stock and options markets earn 39%. In contrast, those that trade only in options earn a staggering average return of 353%. The return calculation for stocks does not reflect any leverage that traders might be using. For insider trades in stocks to earn a similar rate of return as in options, they must leverage their returns by a factor of 15 (353%/24%, which is approximately the average omega for our sample of options).

We examine how insider trading affects the information reflected in prices (i.e., how quotes respond to informed trading) and whether price discovery measures can detect the presence of insider trading. We have a sample of news announcements (the information on which the insiders traded) for 36 stocks, with all necessary data (including 2,031 stock-days, of which 212 are stock-days known to have illegal insider trading).[22] If the price discovery metrics detect the

[21] Several studies find that the sample selection bias that arises from examining detected insider trading is small. For example, Meulbroek (1992) finds similar results for cases initiated by public complaints and cases initiated by exchange referrals.

[22] To allow for variation per stock, our sample is constructed using stocks that have listed options and episodes of insider trading that involve at least six days on which insiders illegally trade in either market. Trades made by insiders to close out their positions are considered to be uninformed. In the subsequent analysis, we do not classify these trades as illegal insider trades because once the private information held by an insider is revealed to the market

presence of insider trading, we would expect higher options price discovery shares when insiders trade in the options market only, compared to stock-days in which insiders trade in the stock market only, or in which insiders do not trade.

To test whether the price discovery measures reflect informed trading, we regress them on three dummy variables: (1) $OptionDV_{it}$, which takes the value of one if on that stock-day insider trading occurs only in options; (2) $StockDV_{it}$, which takes the value of one if insider trading occurs only in stocks; and (3) $BothDV_{it}$, which takes the value of one if insider trading occurs in both stocks and options. For stock-days with no known insider trading, all three of the dummy variables are zero. We also include stock fixed effects and a time trend ($Trend_t$).

Table 5 reports the regression results, using *IS*, *ILS*, and *ILI* as the price discovery measures in Models 1, 2, and 3, respectively. The results show that the options *IS* is *lower* when informed trading occurs in the options market (negative coefficient of $OptionDV_{it}$) contrary to the expectation that these price discovery measures are able to detect the presence of informed trading. In contrast, Models 2 and 3 show that the options *ILS* and *ILI* are significantly *higher* when insiders trade in options; *ILS* is 7.12 percentage points higher than at other times and *ILI* is 6.14 percentage points higher, holding other factors constant.[23] The regressions indicate that *ILS* and *ILI*, but not *IS*, reveal the presence of insider trading. These findings support the use of *ILS* and *ILI* as measures of where fundamental information is first impounded into prices.

< Table 5 here >

---

they no longer have an informational advantage. For each insider trading prosecution, we obtain data for that stock, starting one month before the first insider trade and ending one month after the last insider trade.

[23] These results are also robust when controlling for stock volatility. In all three models, we observe the expected negative coefficient on $StockDV_{it}$. The regressions in which *ILI* is the dependent variable are effectively "linear probability models." We obtain similar results using probit regression models, which we report in the Internet Appendix (see Tables IA.5, IA.7, IA.8, and IA.9). We also obtain similar results when we apply logit transformations to convert *IS* and *ILS* into unbounded variables (see Table IA.5, IA.7, and IA.8 of the Internet Appendix).

There are potential limitations in using the insider trading cases to validate the price discovery measures. First, the analysis is a test of joint hypotheses: (i) that the insider trades are recognized by the market as informed and consequently that insider information is impounded into prices, and (ii) that the price discovery measures can detect the market in which the information was first reflected in prices. In some cases, prices do not reveal the presence of informed traders (e.g., Collin-Dufresne and Fos, 2015), therefore weakening the expected relation between insider trading and price discovery measures. Second, the insider trades might account for only a small fraction of the total trading and thus the insiders' impact on prices could be small relative to other factors. Third, we observe only *detected* insider trades that resulted in legal action by the SEC. There could be other informed trades, possibly based on the same information, occurring in markets other than where the prosecuted insiders traded. It is also possible that because options are considerably less liquid than stocks, it is easier for the SEC to detect insider trading in options by observing unusual options trading prior to price-sensitive announcements. The insider trading detection rate in options might also be higher due to options market makers filing complaints with the SEC when they incur significant losses to informed traders prior to price-sensitive announcements. For these reasons, the share of *prosecuted* illegal insider trading that occurs in options cannot be directly compared to the options share of price discovery.

4.4. Options price discovery around information events

The previous subsection shows that informed trading occurs in options and their trading is reflected in options prices (i.e., it impacts *ILS* and *ILI*). Given that informed investors trade at times when they have an informational advantage, such as when new information is generated or announced to the market, it is natural to ask whether the role of options in price discovery is different around information events than at other times. We address this question using two proxies for situations where significant new information enters the market: (1) stock-days when insider trading occurs, and (2) stock-days on which price-sensitive news is released. We obtain news releases from Thomson Reuters News Analytics (TRNA). We define a stock-day as having a price-sensitive news release if: (i) the number of news items on the given day mentioning the stock exceeds the median number of news items per stock-year during our sample period, and (ii) the average *TRNA* relevance score for the news on a given stock-day exceeds 0.75.[24]

Panel A of Table 6 reports that the mean options *ILI* is 41.04% on days when insiders trade on their private information (this estimate is from the sample of 36 stocks with prosecuted insider trading, which generally have *less* active options markets). Using a difference of means t-statistic, this *ILI* estimate is 12.10 percentage points higher than the average options *ILI* for stock-days with no insider trading in our sample of 35 large stocks. Similarly, options *ILS* is 14.18 percentage points higher on insider trading stock-days compared to other stock-days. The differences in means are statistically significant and large in magnitude, particularly relative to the pooled sample means of those measures.

< Table 6 here >

[24] Important news is typically reported across a number of media outlets and therefore criterion (i) helps identify important news, avoiding casual mentions of the stock. Criterion (ii) ensures the news is specific to the company, which further helps avoid casual mentions of stocks in general market news. TRNA gives each news announcement: (a) a discrete sentiment score if it contains good news, no-news, or bad news, and (b) a continuous relevance score ranging between zero and one determined by how many times the company is mentioned relative to other companies in the news announcement. Our results are robust to alternative definitions of price-sensitive news including different thresholds ranging from 0.50 to 0.90 of the average relevance score, considering good news only, and considering bad news only.

Panel B of Table 6 shows that the options share of price discovery also increases around price-sensitive news releases. The mean options *ILI* on stock-days with price-sensitive news is 32.04%, which is a statistically significant 3.33 percentage points higher than stock-days with no price-sensitive news. Similarly, *ILS* is 2.33 percentage points higher on news days. The post-announcement return drift reported in previous studies suggests that it takes time for investors to process the information in public announcements. It is likely that some traders are faster (or more accurate) than others in processing public news and our results suggest that a proportion of such traders use the options market to maximize their returns.

The increase in options *ILI* and *ILS* around news releases is not due to an increase in sampling variation. The 95% confidence intervals for both mean *ILS* and mean *ILI* (generated from the bootstrap distribution) are in fact *narrower* on news stock-days than at other times, pointing away from an increase in sampling variation as the driver of the increased values. The differences in the average widths of the 95% confidence intervals on news stock-days compared to other stock-days are significantly narrower, with *t*-statistics of -8.50 and -6.28 for *ILS* and *ILI*, respectively.

Using multivariate regressions, we confirm that the increased options share of price discovery around price-sensitive news announcements is robust to various controls. We regress the price discovery measures on a dummy variable for the release of price-sensitive news ($News_{it}$), controlling for stock fixed effects, DV_t^{GF}, DV_t^{OMM}, and a time trend. For example, Table 7 reports that options *ILI* is 4.19 percentage points higher around price-sensitive news, holding other factors fixed.

< Table 7 here >

Our finding that the options market's share of price discovery is higher during periods of insider trading and price-sensitive news releases reinforces the importance of options in impounding new information. Higher options *ILI* and *ILS* during such periods indicate that options are often the first of the two markets to reflect news or the price impact of informed traders. Our findings support the interpretation that the abnormal trading activity previously documented in the options market around information events is, at least in part, due to informed trading.

4.5. The determinants of price discovery in the options market

Our results suggest that options play an important, rather than a negligible, role in price discovery, especially around information events, at least in part because informed traders utilize options. In this subsection, we investigate *why* informed traders choose to trade options despite their relative illiquidity, and *how* their trading affects the options market. Given that *ILI* identifies which market is the informational leader and is able to detect the presence of informed trading, we use the options *ILI* as our main measure of the relative amount of informed trading in options. We examine how informed traders' choice of market is affected by relative liquidity, volatility, and leverage. We start with OLS panel regressions before estimating two-stage least squares (2SLS) instrumental variables (IV) models to address bi-directional causality; namely, informed traders might choose to trade in options due to particular market conditions, but their choice of market can itself influence those conditions.

The liquidity hypothesis predicts that informed traders will choose the most liquid market to minimize the price impact of their trades and thus maximize the value of their information.

Fleming et al. (1996) find support for this hypothesis in stock, futures, and options markets. Under the liquidity hypothesis, we would expect a negative (positive) relation between options price discovery shares and options bid-ask spreads (options trading volume).

The uncertainty hypothesis stems from Capelle-Blancard (2001), who models the strategic interaction between traders informed about the direction of future stock price movements (directional-informed traders) and traders informed about volatility (volatility traders). The presence of volatility traders increases options spreads, shifting directional-informed traders from the options market to the stock market. Under this hypothesis, uncertainty (proxied by stock price volatility) should be negatively related to the amount of price discovery occurring in the options market.

Chakravarty et al. (2004) examine the competing determinants of price discovery in stock and options markets. For comparability, we replicate their regressions, using IS to measure price discovery and similar explanatory variables, before identifying the determinants of ILI using an expanded set of determinants.

We estimate panel regressions using stock-day observations:

$$PD_{it} = \alpha_i + \beta_1 Spread_{it} + \beta_2 Volume_{it} + \beta_3 Volatility_{it} + \sum_j \gamma_j \, OtherDeterminants_{j,it} + \varepsilon_{it}, \quad (11)$$

where PD_{it} is the price discovery measure for the options market, $Spread_{it}$ is the ratio of the time-weighted average quoted bid-ask spread in the options market (averaged across put-call pairs) to that of the stock market, $Volume_{it}$ is the ratio of options omega-adjusted dollar volume (summed across put-call pairs) to stock market traded dollar volume, and $Volatility_{it}$ is the standard deviation of one-minute stock midquote returns. All explanatory variables are in natural log form.

Table 8 reports the regression results. In Model 1, we find a negative relation between the options market *IS* and options bid-ask spreads. This result is qualitatively similar to Chakravarty et al. (2004), who interpret it as support for the liquidity hypothesis—wider options spreads discourage informed trading in the options market, which reduces the options market's contribution to price discovery. However, given the evidence that *IS* is biased downward by noise (Yan and Zivot, 2010; Putniņš, 2013; and the simulations reported in Table 2), an alternative explanation for this result is that wider options spreads are associated with more noise in options prices and thus lower *IS* due to an increased downward bias. Consistent with this alternative interpretation, options *ILI* and options spreads in Model 3 are positively related (the same is true for *ILS* in Model 2) and highly statistically significant. Economically, Model 3 implies that a 1% increase in the options relative spread is associated with an increase in *ILI* that is equivalent to 0.40% of its mean.[25]

< Table 8 here >

The positive relation between the options bid-ask spreads and options' contribution to price discovery is consistent with an adverse selection mechanism. Higher levels of informed trading in the options market increase adverse selection risk for options market makers, leading to wider options spreads (e.g., Glosten and Milgrom, 1985), yet at the same time the increased informed trading in options results in more price discovery in the options market. This finding is also consistent with Hu (2014), who shows that option market makers commonly delta-hedge option

[25] In Model 3, the dependent variable is ILI_{it} and the independent variable is $\ln\left(\frac{Spread_{option,it}}{Spread_{stock,it}}\right)$. A 1% increase in $\left(\frac{Spread_{option,it}}{Spread_{stock,it}}\right)$ is associated with an increase in ILI_{it} of $11.74\ \ln(1.01) = 0.1168$ (with *ILI* measured on a scale of 0-100), with that increase being 0.40% of the pooled sample mean of *ILI* (28.94%).

trades in the stock market and the stock market order imbalance from delta-hedging is informative about future stock returns.

Models 2 and 3 also show a significant and positive relation between the relative volume in options and options price discovery. This is consistent with a pooling equilibrium in which informed traders disguise their trades among those of uninformed traders (Chowdhry and Nanda, 1991; Easley et al., 1998).

The coefficient of $Volatility_{it}$ is positive in all specifications with ILI as the dependent variable, providing no support for the uncertainty hypothesis. The relation between ILI and $Volatility_{it}$ is consistent with more price discovery occurring in options on more volatile days, such as those where insiders trade (see Table 5) and on days in which price-sensitive news is released (see Table 7).

In Models 4 to 6, we use ILI as the dependent variable and examine additional determinants of information leadership. Theory suggests that informed traders will be attracted by the leverage in options. Our model-free approach to obtaining options-implied stock prices means that we can only calculate price discovery for put-call pairs with the same strike price, not individual put and call options. In a given pair, the price (and thus leverage) of the put option will usually be different to the price (and leverage) of the call option. Therefore, we develop a measure of leverage that is applicable to put-call pairs by considering whether the put or the call is more likely to be traded by an informed trader:

$$Leverage_{it} = LeverageCall_{it}\, \mathbf{1}_{it}^{\{r>0\}} + LeveragePut_{it}\, \mathbf{1}_{it}^{\{r<0\}}, \tag{12}$$

where $LeverageCall_{it}$ and $LeveragePut_{it}$ is the omega of each option (the absolute delta multiplied by the ratio of the stock price to the option price) and $\mathbf{1}_{it}^{\{r>0\}}$ and $\mathbf{1}_{it}^{\{r<0\}}$ are indicator variables that equal one if the daily stock return at $t+1$ is positive or negative, respectively. If the daily stock return is positive, $Leverage_{it}$ reflects the leverage in the call option, and vice versa. This measure is based on the assumption that informed traders with good (bad) news will buy call (put) options rather than sell put (call) options. This assumption is supported by the data: in 155 of the 157 announcements in which insiders illegally trade in options, insiders buy call or put options as their first trade. The direction of the stock's return is a proxy for whether informed trader(s) have good or bad news. For each stock-day, we compute the average of the leverage variable for the put-call pairs used in the estimation of the options price discovery measures. Under the leverage hypothesis, we expect a positive relation between leverage and options share of price discovery.

Model 4 shows that options contribute more to price discovery when they have higher leverage.[26] This result is consistent with theoretical predictions made by Easley et al. (1998), which suggest that informed traders are attracted to the options market by the ability to leverage their returns.

We also include a measure of the relative quoting activity, $Quotes_{it}$, defined as the ratio of the number of NBBO quote changes in the options market (summing across put-call pairs) to that of the stock market. From the earlier descriptive statistics, we know that the average number of quote changes per stock-day in options is around a quarter of the corresponding number in stocks. Quoting activity in options is likely to increase with the amount of trading activity in options (the regression accounts for this by controlling for $Volume_{it}$), but also with greater use of autoquotation algorithms by options market makers. Greater use of autoquotation algorithms might increase the tendency for options to follow stocks and therefore more frequent quote

[26] Our results are also robust to using different definitions of leverage including the average of the call and put option omega, and the maximum of the call or put option omega. All of these estimates are likely to understate the effect of leverage because in jointly considering price discovery across many put-call pairs, they take the average leverage across multiple pairs of options, which does not precisely identify the leverage of the particular options that were the first to reflect new information (a form of measurement error).

changes in options might be associated with decreased price discovery. Model 4 shows that more frequent quote changes in options (holding options trading volume fixed) are associated with less price discovery occurring in options. This result is consistent with our conjecture that greater use of autoquotation algorithms increases both quoting activity in options and thereby the tendency for options to mechanically follow stocks.

Several factors determine option leverage, either by affecting the option price relative to the stock price or by affecting the option delta. Some of these factors may also affect where informed traders choose to trade and thus price discovery for reasons other than leverage. For example, informed traders might have a preference for particular maturities, selecting options that have expiry just beyond the expected horizon of their information. We test for this using a series of dummy variables, DV_{it}^{a-b}, equal to one if the time to expiry is between a and b days, inclusive (options with expiry between 40 and 70 days are the base case). We also examine the implied volatility from the Black-Scholes model ($ImpVol_{it}$), which affects leverage through the options price level. Informed traders in the options market might also prefer particular strike prices relative to the current stock price—the degree of option "moneyness." We calculate the absolute difference between the underlying stock price and strike price for each put-call pair and then average across put-call pairs for each stock-day ($StrikeDistance_{it}$). An increase in $StrikeDistance_{it}$ decreases the moneyness of the call option (if the stock price is less than the strike price) or the put option (if the stock price is greater than the strike price). Moneyness and leverage are related, but are not identical. Figure IA.4 of the Internet Appendix shows that as the stock price increases relative to the strike price, the leverage of put options increases and the leverage of call options decreases. The changes in $StrikeDistance_{it}$ are linear, while the changes in leverage are non-linear.

In Model 5 in Table 8, we replace the $Leverage_{it}$ variable with its components: strike distance, implied volatility, and time to maturity. Strike distance is not statistically significant but the sign of the coefficient indicates options that are deeper in or out of the money make a larger contribution to price discovery than at-the-money options. Implied volatility is negatively related to options price discovery, consistent with a leverage effect—higher implied volatility means higher option prices and thus less leverage. The time to maturity dummy variables indicate a monotonic tendency for options that are closer to expiry to play a larger role in price discovery. This result is consistent with the notion that informed traders in options have fairly short-horizon information (such as knowledge of an upcoming announcement) and therefore prefer short-dated options, which are cheaper and have higher leverage. Finally, Model 6 shows that the results on the determinants of options price discovery are robust to including stock fixed effects and additional control variables (DV_t^{GF}, DV_t^{OMM}, and a time trend).

The endogeneity of options market liquidity presents a potential complication; informed traders might choose options when they are relatively liquid, but their trades in options impose adverse selection risks on options market makers, which can reduce liquidity (e.g., Easley et al., 1998). To disentangle these effects, we exploit the exogenous reduction in options tick sizes as an instrument for options market liquidity in a 2SLS IV regression framework. The change in tick size is expected to decrease options bid-ask spreads by removing a binding constraint on the width of the spread. Our measure of relative liquidity is $Spread_{it}$, the ratio of the time-weighted average quoted bid-ask spread in the options market to that of the stock market. Under the Penny Pilot program, tick sizes for OPRA options exchanges were reduced from \$0.05 (\$0.10) to \$0.01 (\$0.05) for options with prices less than (greater than) \$3. The reduction in tick sizes occurred in a staggered manner on different dates for different options classes. In total, 29 of 35 stocks in our sample had a reduction in tick size on seven separate dates between February 2007 and August 2010 (six of 35 stocks changed ticker, merged, or delisted prior to the reduction in tick sizes). Our instrument is a dummy variable that takes the value of one from the time the options tick size is reduced (${DV_{it}}^{Tick}$).

To be a valid instrument, the reduction in options tick sizes should not directly affect options price discovery other than through changes in relative liquidity. While this is likely to be true when using *ILI* as the measure of price discovery, it may be violated when using other price discovery metrics due to their sensitivity to noise. For this reason, we focus on *ILI*.[27]

In the first stage, we regress $Spread_{it}$ on the instrumental variable (IV) and a set of control variables including stock fixed effects:

$$Spread_{it} = \alpha_i + \beta {DV_{it}}^{Tick} + \sum_j \gamma_j \, Controls_{j,it} + \varepsilon_{it}. \tag{13}$$

In the second stage, we regress the options *ILI* on fitted values of relative bid-ask spreads ($\widehat{Spread_{it}}$) obtained from the first stage and the same set of control variables (including stock fixed effects):

$$ILI_{it} = \alpha_i + \beta \widehat{Spread_{it}} + \sum_j \gamma_j \, Controls_{j,it} + \varepsilon_{it}. \tag{14}$$

The control variables include $Volume_{it}$, $Quotes_{it}$, $Volatility_{it}$, $Leverage_{it}$, DV_{it}^{GF}, DV_{it}^{OMM}, and $Trend_t$.

The first-stage regression results in Models 1 and 2 of Table 9 show a significant reduction in options bid-ask spreads (relative to stock market spreads) after the reduction in options tick sizes.[28] Models 3 and 4 report results from the second stage. The relation between relative options bid-ask spreads ($\widehat{Spread_{it}}$) and options price discovery is statistically indistinguishable

[27] Consistent with this hypothesis, if we instead measure options price discovery using *IS* (which is sensitive to noise) in our 2SLS IV approach we find similar results to Model 1 in Table 8.
[28] The *F*-statistics for our instrumental variable exceed one, indicating the instrument is strong (Bound et al., 1995).

from zero in both specifications. In contrast to Chakravarty et al. (2004), we conclude that liquidity has an insignificant bearing on informed traders' choice of market. Our findings differ due to using *ILI* rather than *IS* (thereby avoiding measuring a mechanical relation that occurs through the effects of spreads on noise), as well as using an instrumental variable to overcome the endogeneity issue. The coefficient on relative spreads falls from a maximum of 11.74 in the single-stage OLS models to minimum of 3.22 in the IV regressions. Together, the one-stage OLS and 2SLS IV results suggest that informed trading in the options market increases options bid-ask spreads (e.g., due to adverse selection), but relative liquidity (proxied by the bid-ask spread) is not an important determinant of informed traders' choice of market.

< Table 9 here >

4.6. Options price discovery and price disagreements

The analysis so far is based on the VECM approach to estimating price discovery shares, which is common in the market microstructure literature. An alternative approach introduced by Muravyev et al. (2013) is to examine instances of disagreement in the prices of the two markets and the subsequent price movements in each of the markets to resolve the disagreement. A disagreement is when the bid-ask range of the stock market does not overlap with the implied bid-ask range in the options market and the price differential is sufficiently large to imply the presence of an arbitrage opportunity. In this subsection, we apply the price disagreement approach and compare it to the price discovery shares estimated from the VECM approach.

A price disagreement event is deemed to have occurred if either of the conditions in equation (15) or (16) is satisfied:

$$IP_t > P_t: Implied\ Bid_t(K,T) - \ S_t{}^{Ask} \ \geq \$0.02, \text{ and } \frac{Implied\ Bid_t(K,T) - S_t{}^{Ask}}{S_t{}^{Ask}} \geq 0.05\%, \tag{15}$$

$$IP_t < P_t: S_t{}^{Bid} - Implied\ Ask_t(K,T) \geq \$0.02, \text{ and } \frac{S_t{}^{Bid} - Implied\ Ask_t(K,T)}{S_t{}^{Bid}} \geq 0.05\%, \tag{16}$$

where $IP_t > P_t$ is a disagreement event in which the options-implied-bid price is greater than the stock ask price, $IP_t < P_t$ is a disagreement event in which the options-implied-ask price is less than the stock bid price, $S_t{}^{Bid}$ is the stock bid price at time t, and $S_t{}^{Ask}$ is the stock ask price. The conditions identify instances in which there is a relatively large disagreement between the two markets, and an arbitrage opportunity of at least two cents and 5 bps.

In the sample of put-call-pairs used in the earlier sections, we identify 904,430 price disagreement events, made up of 446,877 $IP_t < P_t$ events and 457,553 $IP_t > P_t$ events. The time series of the frequency of price disagreements is correlated with the Market Volatility Index (VIX). Prior to 2007, there is an average (median) of one (zero) disagreement per stock-day. This number increases to 27 (three) by 2009, before falling to seven (two) disagreements by 2013.

The price adjustments in each market following disagreement events contain information about the price discovery process. For example, if stocks always drive price discovery during disagreement events, we would expect to find that subsequent to a disagreement emerging, options prices make 100% of the adjustment to correct the disagreement, effectively "catching up" or following the stock price. Conversely, if options always drive price discovery during disagreement events, stock prices would adjust to correct the disagreement. Finally, if both markets contribute to price discovery during disagreement events, on average, some of the price adjustment to correct the disagreements will occur in options prices and some will occur in stock prices.

We therefore examine each market's price adjustment following disagreement events using the following VECM:

$$\Delta p_t = \alpha Z_{t-1} + \sum_{i=1}^{20} b_i \Delta p_{t-i} + \gamma Dis_{t-1} + e_t, \tag{17}$$

$$Dis_{1t} = \begin{cases} 0 & if \quad no\ disagreement \\ -1 & if \quad IP_t > P_t\ (options\ too\ expensive) \\ +1 & if \quad IP_t < P_t\ (options\ too\ cheap) \end{cases} \tag{18}$$

$$Dis_{2t} = \begin{cases} 0 & if \quad no\ disagreement \\ +1 & if \quad IP_t > P_t\ (options\ too\ expensive) \\ -1 & if \quad IP_t < P_t\ (options\ too\ cheap) \end{cases} \tag{19}$$

where Δp_t is the 2×1 vector of midquote returns (from options-implied prices and stock prices) at a ten-second frequency, Z_{t-1} is an error correction term, Dis_t is the 2×1 vector of disagreement indicators (Dis_{1t} in the options price equation and Dis_{2t} in the stock price equation), and γ captures how much each market moves to correct a disagreement. We sample prices at a lower frequency here (ten seconds as opposed to one second in the rest of the paper) because disagreement events typically last more than a second.[29] We estimate the VECM for each stock in each year, omitting overnight returns.[30]

The adjustment in prices following disagreement events is the sum of two parts: (1) adjustment implied by the error correction term (captured by α), and (2) additional adjustment due to the disagreement event (captured by γ). To interpret the results, in particular the magnitudes of the adjustment coefficients, it is useful to compare them to the typical disagreement size during our sample period, which is 7 bps.

The results from estimating the VECM (equation (17)) indicate that for the full sample, the average price adjustment in the options market in the ten seconds following a typical 7 bps

[29] For example, Muravyev et al. (2013) find that in their sample the median disagreement event durations are 16.4 and 18.3 seconds for the $P_t > IP_t$ and $IP_t > P_t$ events, respectively.
[30] We do this by normalizing the opening stock price to be equal to the previous closing price.

disagreement is 3.35 bps, which is statistically significant at the 5% level (see Table C1 in Appendix C). This adjustment is made up of 1.17 bps implied by the error correction term [i.e., α multiplied by the typical disagreement size (α_{Dis})] and a further 2.18 bps implied by the coefficient γ. Therefore, the estimated total adjustment in options prices in the ten seconds following a disagreement is approximately half (47.86%) of the typical disagreement size.

Repeating the same calculations for stock prices, we find that the average adjustment of stock prices in the ten seconds following disagreements is 0.51 bps, which is statistically significant at the 5% level. This is made up of 0.07 bps implied by the error correction term (α_{Dis}) and an additional 0.44 bps due to the disagreement event (γ). Therefore the estimated total adjustment in stock prices in the ten seconds following a disagreement is approximately 7.29% of the typical disagreement size. These results indicate that the majority of the adjustment in prices following disagreement events occurs in options, consistent with our overall finding that stocks make a larger contribution to price discovery than options. The results also indicate that some of the adjustment following disagreement events occurs in stock prices, consistent with the notion that options drive some of the price discovery during disagreement events. We would not necessarily expect to find options accounting for the same share of price discovery during disagreement events and during other times, in particular considering that disagreement events represent about 1% of the total sample by time. We find similar results using an alternative VECM specification that omits the error correction term and when using a lower (20 second) sampling frequency.

To better understand the relation between the two approaches (the VECM-based price discovery shares and the disagreement approach), we examine the extent to which the post-disagreement price adjustment magnitudes correlate with the price discovery shares in the pooled sample. First, using the estimates of how much stock and options prices adjust in the ten seconds following a disagreement event, we construct a new variable, $Adjust_{it}$, which is the ratio of the adjustment in stock prices to the adjustment in options prices. Higher values of $Adjust_{it}$ indicate greater adjustments in stock prices following disagreement events, relative to those of options prices, which is likely to occur when options contribute more to price discovery. We then regress each of the options price discovery shares (*IS*, *ILS*, and *ILI*) on $Adjust_{it}$, controlling for a number of other variables including $Spread_{it}$, $Volume_{it}$, $Volatility_{it}$, $Quotes_{it}$, $Leverage_{it}$, DV_{it}^{GF}, DV_{it}^{OMM}, and $Trend_t$.

Table 10 reports the results of the regressions. Models 2 and 3 show that options price discovery shares (*ILS* and *ILI*) are positively related to $Adjust_{it}$ controlling for other factors. These results support the notion that when the relative adjustment in stock prices is larger following price disagreements, consistent with options making a larger contribution to price discovery, the options price discovery shares also tend to be higher. Therefore, the two approaches tend to agree about both the variation in options price discovery, and when options make a relatively larger or smaller contribution to price discovery.

< Table 10 here >

Finally, we consider the similarities and differences in how the two approaches analyze price discovery. The results from the two approaches *qualitatively* agree on several aspects, but produce different quantitative estimates. For example, the estimates of both approaches are positively correlated, suggesting that they tend to agree as to when options make a relatively larger or smaller contribution to price discovery. Also, both approaches agree that stocks more often lead in reflecting new information, but that options do contribute to price discovery. Where they differ is in the quantitative estimates of how much options contribute to price discovery. For example, in our sample options *ILI* estimates that options lead stocks approximately 29% of the time, while the analysis of price disagreements suggest that stocks are responsible for about 13% of the adjustment following disagreement events (recall that stock price adjustment is consistent with options contributing to price discovery).[31]

There are a few possible reasons we may observe differences in the estimates from the two approaches. First, *ILI* estimates lead-lag relations between the markets using intraday returns throughout the whole day, whereas the disagreements approach focuses on particular points in time. For example, there are approximately 900,000 price disagreements in our sample with a mean duration of around 15 seconds each, implying that disagreements span a total of approximately 13.5 million seconds in our sample. There are approximately 1.2 billion seconds in our sample in total, implying that options and stock prices disagree about 1% of the time. The estimates suggest that when stock and options prices disagree (i.e., 1% of the time), options contribute less to price discovery (around 13%) than they do at other times. During other times (i.e., 99% of the time), *ILI* suggests that options lead price discovery 29% of the time. The differences in the points in time considered by the two approaches is one way of reconciling the estimates from the two approaches—price discovery contributions vary through time (even intraday) and the two approaches measure price discovery at different points in time.

A second factor contributing to the differences in results is that the two approaches consider different sets of price movements. The VECM-based price discovery shares explicitly separate permanent and temporary price movements. The permanent movements are assumed to reflect

[31] This share of the adjustment is calculated as the ratio of the estimated stock price adjustment (0.51 bps) to the sum of the stock price adjustment and options price adjustment (0.51 bps + 3.35 bps).

new information entering prices, while the temporary movements can be caused by various sources of noise including large uninformed trades, price pressure, pricing errors due to the discrete price grid and so on. Price discovery in the VECM approach is measured using only the permanent (information driven) price movements. In contrast, the disagreements approach has the advantage of being non-parametric, but it does not explicitly separate temporary from permanent price movements. Some of the price disagreements could be triggered by temporary price changes, caused by strong price pressure or large uninformed orders. Therefore, the two approaches also differ in terms of which price movements are counted when estimating which market leads the process of price discovery, which could also contribute to the differences in estimates.

Given that both approaches bring additional information about the price discovery process, when used together they provide a more complete picture of the nature of price discovery. The price disagreement method provides a less parametric approach and focuses on profitable opportunities for market participants. However, as price disagreements do not occur often, in particular during periods of low volatility, the VECM approach can be used to estimate price discovery more generally, including during less volatile periods and at more granular units of analysis, such as individual stock-days. Future work might try and capitalize on the advantages of both approaches by combining them in a way that gets “the best of both worlds.”

5. Conclusion

By disaggregating price discovery into two components—the relative speed with which information is reflected in prices and relative noise—and using a unique dataset of insider trading prosecutions, this paper brings new evidence to the debate about the role of options in price discovery. We find that actively-traded options listed on a sample of large U.S. stocks are the first to reflect new information approximately one-quarter of the time, indicating that they play an important role in price discovery. Using unique data on prosecuted insider trading, we find that around one-third of the time prosecuted insiders trade in options, and when they do, the options information leadership indicator (our main measure of options price discovery) is significantly elevated. Consistent with theoretical predictions, we show that leverage is one of the key factors that attract informed traders to the options market.

In contrast, traditional price discovery measures do not increase in response to insider trading and underestimate the amount of price discovery that occurs in options. Our findings indicate that the low options information shares reported in previous studies (as low as 6% in recent studies) underestimate the share of price discovery that occurs in options due to their relative illiquidity. The low information shares indicate that options prices are noisier than stock prices, *not* that they are unimportant in impounding new information. Our Monte Carlo simulations and empirical proxies of relative noise in options prices support these conclusions.

We also find that options contribute more to price discovery at times when information enters the market, such as when insiders trade on advance knowledge of corporate announcements and when price-sensitive news is released. Because informed trading creates adverse selection risk, options spreads tend to be wider at times when options contribute more to price discovery. Our results suggest that the abnormal trading characteristics in options around various types of news (reported in a growing number of studies) are at least in part due to informed trading.

Our findings have a number of practical implications. Illegal insider trading is associated with substantial costs, including (i) the substantial regulatory resources used in lawmaking, monitoring, and enforcement of insider trading rules [e.g., in 2014 the SEC spent approximately $531 million in combating insider trading (SEC, 2013)], and (ii) negative effects on financial markets such as damaged investor confidence in the fairness of markets, which can reduce investor participation and harm liquidity. A better understanding of insider trading strategies can improve the efficiency with which regulatory resources are used. Our findings contribute to this understanding by providing insights about where informed traders choose to trade, what influences their decisions, and how their trading impacts markets. Given the significant share of price discovery occurring in options, regulators should not ignore options markets in their surveillance activities.

References

Ahern, K. R., (2017). Information networks: Evidence from illegal insider trading tips. *Journal of Financial Economics* 125, 26–47.

Amin, K., Lee, C., (1997). Option trading, price discovery, and earnings news dissemination. *Contemporary Accounting Research* 14, 153–192.

Anthony, J. H., (1988). The interrelation of stock and option market trading-volume data. *Journal of Finance* 43, 949–961.

Acharya, V. V., Johnson, T. C., (2010). More insiders, more insider trading: Evidence from private-equity buyouts. *Journal of Financial Economics* 98, 500–523.

Augustin, P., Brenner, M., Subrahmanyam, M. G., (2014). Informed options trading prior to M&A announcements: Insider trading? *Working paper, McGill University and New York University.*

Augustin, P., Brenner, M., Hu, J., Subrahmanyam, M. G, (2015). Are corporate spin-offs prone to insider trading? *Working paper, McGill University, New York University and Singapore Management University.*

Back, K., (1993). Asymmetric information and options. *Review of Financial Studies* 6, 435–472.

Baillie, R. T., Booth, G. G., Tse, Y., Zabotina, T., (2002). Price discovery and common factor models. *Journal of Financial Markets* 5, 309–321.

Bali, T. G., Hovakimian, A., (2009). Volatility spreads and expected stock returns. *Management Science* 55, 1797–1812.

Barraclough, K., Robinson, D. T., Smith, T., Whaley, R. E., (2013). Using option prices to infer overpayments and synergies in M&A transactions. *Review of Financial Studies* 26, 695–722.

Berkman, H., McKenzie, M. D., Verwijmeren, P., (2017). Hole in the wall: Informed short selling ahead of private placements. *Review of Finance* 21, 1047–1091.

Black, F., (1975). Fact and fantasy in the use of options. *Financial Analysts Journal* 31, 36–72.

Borochin, P., Golec, J. H., (2016). Using options to measure the full value-effect of an event: Application to Obamacare. *Journal of Financial Economics* 120, 169–193.

Bound, J., Jaeger, D. A., Baker, R. M., (1995). Problems with instrumental variables estimation when the correlation between instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90, 443–450.

Brogaard, J., Hendershott, T., Riordan, R., (2019). Price discovery without trading: Evidence from limit orders. *Journal of Finance* 74, 1621–1658.

Capelle-Blancard, G., (2001). Volatility trading in the option market: How does it affect where informed traders trade? *Working paper, University of Paris.*

Cao, C., Chen, Z., Griffin, J. M., (2005). Informational content of option volume prior to takeovers. *Journal of Business* 78, 1073–1109.

Chakravarty, S., Gulen, H., Mayhew, S., (2004). Informed trading in stock and option markets. *Journal of Finance* 59, 1235–1257.

Chan, K., Ge, L., Lin, T.C., (2015). Informational content of option trading on acquirer announcement return. *Journal of Financial and Quantitative Analysis* 50, 1057–1082.

Chowdhry, B., Nanda, V., (1991). Multimarket trading and market liquidity. *Review of Financial Studies* 4, 483–511.

Collin-Dufresne, P., Fos, V., (2015). Do prices reveal the presence of informed trading? *Journal of Finance* 70, 1555–1582.

Czerwonko, M., Khoury, N., Perrakis, S., Savor, M., (2012). Tick size, microstructure noise, informed trading and volatility inversion effects on price discovery in options markets: Theory and empirical evidence. *Working paper, McGill University, University of Quebec, and Concordia University*.

Easley, D., O'Hara, M., Srinivas, P., (1998). Option volume and stock prices: Evidence on where informed traders trade. *Journal of Finance* 53, 431–465.

Fleming, J., Ostdiek, B., Whaley, R. E., (1996). Trading costs and the relative rates of price discovery in stock, futures and option markets. *Journal of Futures Markets* 16, 353–387.

Ge, L., Tse-Chun, L., Pearson, N., (2016). Why does the option to stock volume ratio predict stock returns? *Journal of Financial Economics* 120, 601–622.

Gharghori, P., Maberly, E. D., Nguyen, A., (2017). Informed trading around stock split announcements: Evidence from the option market. *Journal of Financial and Quantitative Analysis* 52, 705–735.

Glosten, L. R., Milgrom, P. R., (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14, 71–100.

Gonzalo, J., Granger, C., (1995). Estimation of common long-memory components in cointegrated systems, *Journal of Business and Economics Statistics* 13, 27–35.

Grammig, J., Peter, F. J., (2013). Telltale tails: A new approach to estimating unique market information shares. *Journal of Financial and Quantitative Analysis* 48, 459–488.

Hao, Q., (2016). Is there information leakage prior to share repurchase announcements? Evidence from daily options trading. *Journal of Financial Markets* 27, 79–101.

Hasbrouck, J., (1991). The summary informativeness of stock trades: An econometric analysis. *Review of Financial Studies* 4, 571–595.

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.

Hendershott, T., Jones, C. M., Menkveld, A. J., (2011). Does algorithmic trading improve liquidity? *Journal of Finance* 66, 1–33.

Holowczak, R., Simaan, Y., Wu, L., (2006). Price discovery in the US stock and option markets: A portfolio approach. *Review of Derivatives Research* 9, 37–65.

Hu, J., (2014). Does option trading convey stock price information? *Journal of Financial Economics* 111, 625–645.

Hu, J., (2017). Option listing and information asymmetry. *Review of Finance,* forthcoming.

Johnson, T. L., So, E. C., (2012). The option to stock volume ratio and future returns. *Journal of Financial Economics* 106, 262–286.

Kacperczyk, M., Pagnotta, E. S., (2019). Chasing private information. *Review of Financial Studies* (forthcoming).

Lin, T., Lu, X., (2015). Why do options prices predict stock returns? Evidence from analyst tipping. *Journal of Banking and Finance* 52, 17–28.

Lin, T., Lu, X., (2016). Do short-sale costs affect put options trading? Evidence from separating hedging and speculative shorting demands. *Review of Finance* 20, 1911–1943.

Manaster, S., Rendleman Jr. R. J., (1982). Option prices as predictors of equilibrium stock prices. *Journal of Finance* 37, 1043–1057.

Meulbroek, L., (1992). An empirical analysis of illegal insider trading. *Journal of Finance* 47, 1661–1699.

Muravyev, D., Pearson, N. D., Broussard, J. P., (2013). Is there price discovery in equity options? *Journal of Financial Economics* 107, 259–283.

Pan, J., Poteshman, A, M., (2006). The information in option volume for future stock prices. *Review of Financial Studies* 19, 871–908.

Podolski, E. J., Truong, C., Veeraraghavan, M., (2013). Informed options trading prior to takeovers: Does the regulatory environment matter? *Journal of International Financial Markets, Institutions and Money* 27, 286–305.

Putniņš, T. J., (2010). Naked short sales and fails-to-deliver: An overview of clearing and settlement procedures for stock trades in the USA. *Journal of Securities Operations and Custody* 2, 340–350.

Putniņš, T. J., (2013). What do price discovery metrics really measure? *Journal of Empirical Finance* 23, 68–83.

Rourke, T., (2013). Price discovery in near- and away-from-the-money option markets. *Financial Review* 48, 25–48.

SEC, (2013). FY2014 Budget request by program. *US Securities and Exchange Commission*.

Stock, J. H., Watson, M. W., (1988). Testing for common trends. *Journal of the American Statistical Association* 83, 1097–1107.

Yan, B., Zivot, E., (2010). A structural analysis of price discovery measures. *Journal of Financial Markets* 13, 1–19.

Appendix A: Estimation of price discovery shares

From the reduced form VECM estimates (equation (7)), we derive the corresponding infinite lag VMA representation in structural form (assuming recursive contemporaneous causality running from the first through to the fourth price series):

$$\Delta p_{1,t} = \sum_{l=0}^{\infty} A_{1,l}\varepsilon_{1,t-l} + \sum_{l=1}^{\infty} A_{2,l}\varepsilon_{2,t-l} + \sum_{l=1}^{\infty} A_{3,l}\varepsilon_{3,t-l} + \sum_{l=1}^{\infty} A_{4,l}\varepsilon_{4,t-l} \quad \text{(A.1a)}$$

$$\Delta p_{2,t} = \sum_{l=0}^{\infty} B_{1,l}\varepsilon_{1,t-l} + \sum_{l=0}^{\infty} B_{2,l}\varepsilon_{2,t-l} + \sum_{l=1}^{\infty} B_{3,l}\varepsilon_{3,t-l} + \sum_{l=1}^{\infty} B_{4,l}\varepsilon_{4,t-l} \quad \text{(A.1b)}$$

$$\Delta p_{3,t} = \sum_{l=0}^{\infty} C_{1,l}\varepsilon_{1,t-l} + \sum_{l=0}^{\infty} C_{2,l}\varepsilon_{2,t-l} + \sum_{l=0}^{\infty} C_{3,l}\varepsilon_{3,t-l} + \sum_{l=1}^{\infty} C_{4,l}\varepsilon_{4,t-l} \quad \text{(A.1c)}$$

$$\Delta p_{4,t} = \sum_{l=0}^{\infty} D_{1,l}\varepsilon_{1,t-l} + \sum_{l=0}^{\infty} D_{2,l}\varepsilon_{2,t-l} + \sum_{l=0}^{\infty} D_{3,l}\varepsilon_{3,t-l} + \sum_{l=0}^{\infty} D_{4,l}\varepsilon_{4,t-l}\ . \quad \text{(A.1d)}$$

We obtain the structural VMA coefficients by computing the orthogonalized impulse response functions and the (contemporaneously uncorrelated) structural VMA errors ($\varepsilon_{1,t}$ to $\varepsilon_{4,t}$) by mapping their relation to the reduced form errors.

The permanent price impacts of shocks to the four prices are easily obtained from the structural VMA. For example, a unit shock to the first price ($\varepsilon_{1,t} = 1$) has a permanent effect on all of the prices equal to $\theta_{\varepsilon 1} = \sum_{l=0}^{\infty} A_{1,l}$.[32] Similarly, the permanent price impacts of shocks to the second, third, and fourth prices are given by $\theta_{\varepsilon 2} = \sum_{l=1}^{\infty} A_{2,l}$, $\theta_{\varepsilon 3} = \sum_{l=1}^{\infty} A_{3,l}$, and $\theta_{\varepsilon 4} = \sum_{l=1}^{\infty} A_{4,l}$.

In Hasbrouck's (1995) temporary-permanent decomposition, which is based on Stock and Watson's (1988) common trend representation, innovations in the permanent component (the efficient price, m_t) are given by:

$$\Delta m_t = \theta_{\varepsilon 1}\varepsilon_{1,t} + \theta_{\varepsilon 2}\varepsilon_{2,t} + \theta_{\varepsilon 3}\varepsilon_{3,t} + \theta_{\varepsilon 4}\varepsilon_{4,t}. \quad \text{(A.2)}$$

The variance of the innovations in the efficient price is therefore:

[32] The permanent price impact is the same for all prices, $\theta_{\varepsilon 1} = \sum_{l=0}^{\infty} A_{1,l} = \sum_{l=0}^{\infty} B_{1,l} = \sum_{l=0}^{\infty} C_{1,l} = \sum_{l=0}^{\infty} D_{1,l}$, because permanent impacts are innovations in the efficient value and the efficient value is common to all prices as they refer to the same underlying asset (Hasbrouck, 1995).

$$Var(\Delta m_t) = Var\big(\theta_{\varepsilon 1}\varepsilon_{1,t} + \theta_{\varepsilon 2}\varepsilon_{2,t} + \theta_{\varepsilon 3}\varepsilon_{3,t} + \theta_{\varepsilon 4}\varepsilon_{4,t}\big)$$

$$= \theta_{\varepsilon 1}^2 Var\big(\varepsilon_{1,t}\big) + \theta_{\varepsilon 2}^2 Var\big(\varepsilon_{2,t}\big) + \theta_{\varepsilon 3}^2 Var\big(\varepsilon_{3,t}\big) + \theta_{\varepsilon 4}^2 Var\big(\varepsilon_{4,t}\big), \quad \text{(A.3)}$$

because the structural VMA errors are uncorrelated by construction (contemporaneous correlation in the reduced form errors is absorbed in the recursive contemporaneous effects that are part of the structural VMA). Hasbrouck (1995) information shares (*IS*) are then easily obtained as each price's contribution to the variance of the efficient price innovations:

$$IS_1 = \frac{\theta_{\varepsilon 1}^2 Var(\varepsilon_{1,t})}{Var(\Delta m_t)}\ ,\ IS_2 = \frac{\theta_{\varepsilon 2}^2 Var(\varepsilon_{2,t})}{Var(\Delta m_t)},\ IS_3 = \frac{\theta_{\varepsilon 3}^2 Var(\varepsilon_{3,t})}{Var(\Delta m_t)}\ ,\ IS_4 = \frac{\theta_{\varepsilon 4}^2 Var(\varepsilon_{4,t})}{Var(\Delta m_t)}. \quad \text{(A.4)}$$

The Gonzalo and Granger (1995) component shares (*CS*) are obtained by normalizing the permanent price impacts of each price series in the reduced form model:[33]

$$CS_1 = \frac{\theta_{e1}}{\theta_{e1}+\theta_{e2}+\theta_{e3}+\theta_{e4}}, CS_2 = \frac{\theta_{e2}}{\theta_{e1}+\theta_{e2}+\theta_{e3}+\theta_{e4}}, CS_3 = \frac{\theta_{e3}}{\theta_{e1}+\theta_{e2}+\theta_{e3}+\theta_{e4}}, CS_4 = \frac{\theta_{e4}}{\theta_{e1}+\theta_{e2}+\theta_{e3}+\theta_{e4}}.$$

(A.5)

Finally, we calculate the information leadership share (*ILS*) and information leadership indicator (*ILI*) by extending the approach in Yan and Zivot (2010) and Putniņš (2013) to the case of multiple markets. In the two-price case, market i's propensity to reflect new information (how much market i's price responds to an innovation in the efficient price) can be obtained from the ratio $\beta_i = IS_i/CS_i$, which when normalized gives the information leadership share, $ILS_i = \beta_i^2/(\beta_1^2 + \beta_2^2)$.[34] In the four-price case, the information leadership shares are:

$$ILS_1 = \frac{\beta_1^2}{\beta_1^2+\beta_2^2+\beta_3^2+\beta_4^2},\ ILS_2 = \frac{\beta_2^2}{\beta_1^2+\beta_2^2+\beta_3^2+\beta_4^2},\ ILS_3 = \frac{\beta_3^2}{\beta_1^2+\beta_2^2+\beta_3^2+\beta_4^2},\ ILS_4 = \frac{\beta_4^2}{\beta_1^2+\beta_2^2+\beta_3^2+\beta_4^2}.$$

(A.6)

[33] Obtaining the permanent price impacts of the reduced form model is similar to the procedure for the structural model, except that simple impulse response functions from the reduced form VECM are used instead of orthogonalized impulse response functions. We ensure all permanent price impacts are non-negative so that the *CS* take the range [0,1].

[34] β_i can also be equivalently obtained from a regression of market i's price innovations (the structural model errors, $\varepsilon_{i,t}$) on the efficient price innovations (Δm_t): $\varepsilon_{i,t} = \beta_i \Delta m_t + \eta_t$. This provides an interpretation of β_i as the proportion of the efficient price innovation that is immediately reflected in the price of market i.

The information leadership indicator for market i in the two-price case is one if $ILS_i > 0.5$ and zero otherwise. In the four-price case, ILI_i is one if market i has the highest *ILS* of all four markets and zero otherwise:

$$ILI_i = \begin{cases} 1 & \text{if} \quad ILS_i > ILS_k \quad \forall \; k \neq i \\ 0 & \text{otherwise} \end{cases}. \tag{A.7}$$

The options price discovery measure for a given stock-day is obtained by summing the price discovery measures of the valid put-call pairs on that stock-day. For example, if the stock price is p_1 and the three options-implied stock prices are p_2, p_3, and p_4, then the options *IS* is given by $IS_2 + IS_3 + IS_4$. The options *CS*, *ILS*, and *ILI* are obtained in a similar manner.

Hasbrouck's (1995) *IS* are not unique; they depend on the ordering of the prices in the estimation procedure because of the recursive contemporaneous causality assumed to run from the first through to the last price series (this assumption is implicit in procedures that use Cholesky factorization of the covariance matrix of reduced form errors, and explicit in the structural VMA that we present above). We apply the standard approach used in the literature and estimate *IS* (and subsequently *ILS* and *ILI*) for each price under all possible orderings of the prices. This gives a range of *IS* values for each price, with the minimum and maximum values providing the lower and upper bounds on *IS*. Following Baillie et al. (2002) and many subsequent papers, we take the average of the upper and lower bound to obtain a single *IS* estimate for each price.

To obtain standard errors for the price discovery measures, we apply a bootstrap procedure to every stock-day. The procedure recognizes that the price discovery measures are all calculated from the parameter estimates in the reduced form VECM (equation (7)), involving many non-linear transformations of those parameters. Error in estimating the VECM parameters is passed through to the price discovery measures. While the error in the VECM parameter estimates can be quantified with standard approaches, the complexity of the transforms to obtain the price discovery measure makes analytical approaches to calculating their standard errors infeasible. For each stock-day, after estimating the VECM, the bootstrap procedure simulates 100 sets of VECM parameter estimates with random perturbations from the parameter point estimates. The perturbations are drawn from multivariate normal distributions with expected value of zero and variance-covariance obtained from the variance-covariance matrix of the VECM parameter estimates. For each set of perturbed VECM parameter estimates, we calculate the price discovery measures following the procedure described in this Appendix. These steps generate a bootstrap distribution for each price discovery estimate for each stock-day, as well as distributions for quantities such as the mean of the price discovery metrics across multiple stock-days. From the bootstrap distributions we quantify the sampling error, for example, by computing percentiles and using them as confidence intervals for the price discovery estimate, or by computing the standard deviation of the distribution of estimates.

Appendix B: What components of price discovery do the empirical metrics capture?

In this appendix, we describe two sets of analyses that examine the components of price discovery captured by the price discovery measures (*IS*, *ILS*, and *ILI*), with particular focus on how the metrics are affected by noise. The first set of analyses is based on simulated data; the second on empirical data.

B.1. Monte Carlo simulations

Our aim is to simulate stylized models that reflect U.S. stock and options markets. Importantly, to gain insight as to what the price discovery metrics capture, the simulated models have to be such that we unambiguously know (by construction) how each market contributes to price discovery. Following Hasbrouck (2002) and Putniņš (2013), within a given stock-day, the fundamental value of the stock is assumed to follow a random walk,

$$m_t = m_{t-1} + u_t, \qquad u_t \sim N(0, \sigma_u), \tag{B.1}$$

where m_t is the fundamental value at time t, and u_t is i.i.d. normal. We set $\sigma_u = 1$ bps per second similar to the average volatility of U.S. stocks (Hendershott et al., 2011). The stock price at time t, $p_{1,t}$, tracks the fundamental value with a delay of δ_1 periods, and contains noise, σ_{s_1}, where $s_{1,t}$ is i.i.d. normal:

$$p_{1,t} = m_{t-\delta_1} + s_{1,t}, \qquad s_{1,t} \sim N\left(0, \sigma_{s_1}\right). \tag{B.2}$$

Options prices follow the Black-Scholes model. They also track the fundamental value, but with a different delay of δ_2 periods and a different amount of noise, σ_{s_2}:

$$p_{C,t} = m_{t-\delta_2} N(d_1) - Ke^{-rT} N(d_2) + s_{2,t}, \qquad s_{2,t} \sim N\left(0, \sigma_{s_2}\right), \tag{B.3}$$

$$p_{P,t} = Ke^{-rT} N(-d_2) - m_{t-\delta_2} N(-d_1) + s_{3,t}, \qquad s_{3,t} \sim N\left(0, \sigma_{s_2}\right), \tag{B.4}$$

where $p_{C,t}$ is the call price at time t, $p_{P,t}$ is the put price, K is the strike price, r is the continuously compounded risk-free rate, T is the time-to-maturity, and $N(d_i)$ is the probability that a normally distributed variable with a mean of zero and standard deviation of one is less than d_i.[35]

Therefore, in equations (B.2) to (B.4), δ_i characterizes the speed with which market i reflects new information, and σ_{s_i} characterizes the noise in market i (due to illiquidity, the discrete price grid, and so on). An additional source of noise in options prices arises from the conversion of options prices to options-implied stock prices as per the procedure described in Subsection 3.1.

We assume that on any given day, either the options prices are faster to reflect new information ($\delta_1 > \delta_2$), or the stock price is faster ($\delta_1 < \delta_2$). While price discovery shares are unlikely to be fixed over longer periods (e.g., months or years), they are likely to be reasonably stable at intraday horizons, which is why prior studies and also our empirical analysis estimates price discovery shares separately for each day allowing for variation across days. The parameter q, which we vary in our simulations from zero to one in increments of 0.1, is the probability that the options prices are faster to reflect new information on a given day. Thus q is the proportion of the time options lead in price discovery.

Our choice of model parameters is based on empirical features of stock and options markets—our simulations are approximately calibrated to our empirical setting. The cross-correlations of stock and options one-second midquote returns (see Figure 1) shows that options at times lead stocks by up to ten seconds. Consequently, we set the speed differential between

[35] We set K such that the options are initially at-the-money; $r = 1.78\%$, which is the average federal funds rate during our sample period; $T = 2/12$ consistent with our sample selection criteria; and $\sigma = 20.49\%$, which is the average VIX during our sample period. In unreported results, we obtain similar findings when we simulate option prices using away-from-the-money contracts (i.e., where K is set 6% lower and higher than the underlying asset price). Using put-call parity, we calculate the options-implied stock price using the simulated call and put prices.

markets at $\delta_1 = 0$, $\delta_2 = 5$ for days when the stock price is faster to reflect new information (making it five seconds faster) and $\delta_1 = 5$, $\delta_2 = 0$ for days when the options prices are faster. The results are qualitatively similar (and conclusions the same) for larger and smaller differences in the relative speed of the two markets. We fix the noise in the stock price at a moderate level ($\sigma_{s_1} = 5$ bps) and vary the noise in options prices through the range $\sigma_{s_2} \in \{10, \ldots, 25\}$ bps to create variation in the relative noise of options across different simulations. These parameters reflect the empirical fact that options prices are considerably noisier than stocks on average, but that there is variation in the relative noise differences (e.g., options bid-ask spreads are approximately five times larger than stock bid-ask spreads in our sample).[36]

For every one of the 11×6 parameter combinations of q and σ_{s_2}, we simulate 5,000 samples (each corresponding to a day) of 21,600 time series observations (each corresponding to an intraday one-second frequency observation; there are approximately 21,600 seconds in a typical U.S. trading day). We then estimate the price discovery measures (*IS*, *ILS*, and *ILI*) on each sample, and compute the mean of the price discovery measures across the 5,000 samples. Table 2 reports the Monte Carlo simulation results and we discuss the results in Subsection 4.1.

B.2. Empirical measures of noise and their effects on price discovery shares

Turning to the empirical data, we measure the amount of noise in stock and options prices and relate the noise to the ($ILI_{it} - IS_{it}$) differential. Following Gonzalo and Granger (1995), for each one-second interval s, for stock i, on day t, we estimate the common efficient price (CEP_{its}) as a linear combination of the options and stock prices with the component shares as the

[36] We obtain similar results if we vary options noise through an extended range of $\{25, \ldots, 40\}$ bps.

weights associated with each price. Absolute deviations of a put-call pair's implied stock price from the estimated common efficient price are given by:

$$OptionError_{its} = |p_{its} - CEP_{its}| \ . \tag{B.5}$$

For each stock-day, we average $OptionError_{its}$ across the put-call pairs and the intraday one-second intervals to obtain a noise measure of options prices, $OptionNoise_{it}$. Repeating this procedure with the stock price gives $StockNoise_{it}$. We obtain relative measures of options noise by taking the ratio:

$$OptionNoiseRatio_{it} = \frac{OptionNoise_{it}}{OptionNoise_{it} + StockNoise_{it}}. \tag{B.6}$$

As an alternative measure of noise, we calculate the standard deviation of the absolute errors in option prices:

$$OptionStdNoise_{it} = \sqrt{Var(OptionError_{its})}. \tag{B.7}$$

Appendix C: Price disagreements in stock and options markets

Table C1. Price disagreements in stock and options markets

This table reports the estimated adjustment in stock and options prices in the ten seconds following price disagreement events. $DisSize$ is the typical disagreement size measured in bps. α_{Dis} is the adjustment in prices in bps following disagreement events implied by the error correction term [i.e., α in equation (17) multiplied by the typical disagreement size]. γ is the additional adjustment in prices in bps due to the disagreement event, given in equation (17). $\frac{(\alpha_{Dis}+\gamma)}{DisSize}$ is the ratio of the total adjustment in prices to the typical disagreement size expressed as a percentage (range 0–100%). The sample comprises 35 stocks during the April 17, 2003 to April 17, 2013 period.

	$DisSize$	α_{Dis}	γ	$\frac{(\alpha_{Dis}+\gamma)}{DisSize}$
Options	7.00	1.17	2.18	47.86
Stocks	7.00	0.07	0.44	7.29

Figure 1. Cross-correlations of stock and options returns

This figure shows the difference in the mean cross-correlations of stock and options returns for stock-days that have the highest and lowest ILI estimates (the difference is highest ILI stock-days minus lowest ILI stock-days). The cross-correlations (vertical axis) are computed as the correlation of r_t^{STOCK} and r_{t+l}^{OPTION} for each of the lead/lag values (on the horizontal axis) $l = -10, -9, \ldots + 9, +10$, where r_t^{STOCK} is the midquote stock return in the one-second interval t and r_{t+l}^{OPTION} is the average midquote options-implied stock price return (average across put-call pairs) in the one-second interval $t + l$. For example, the cross-correlation difference at lag $l = -1$ of 0.007 (with associated t-statistic of 10) indicates that the correlation between r_t^{STOCK} and r_{t-1}^{OPTION} is 0.007 higher on average for the high-ILI stock-days compared to the low-ILI stock-days and the difference is statistically significant, implying a greater tendency for options returns to lead stock returns when ILI is high. The dotted lines represent 95% confidence levels. The grey bars represent difference of means Satterthwaite t-statistics.

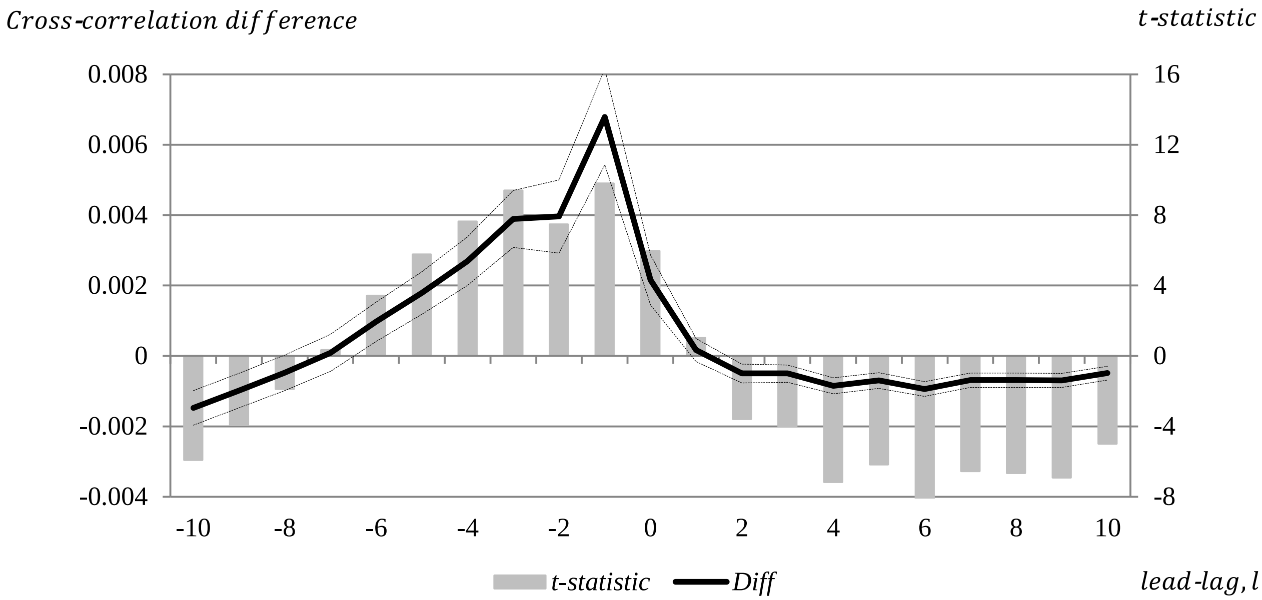

Figure 2. Options market price discovery shares through time

This figure shows the mean options market price discovery shares and mean number of put-call pairs per stock-day through time for our sample of 35 stocks. The price discovery shares are: (i) Hasbrouck information share (IS), (ii) Yan-Zivot-Putniņš information leadership share (ILS), and (iii) information leadership indicator (ILI). All price discovery measures are expressed as percentages (range 0–100%).

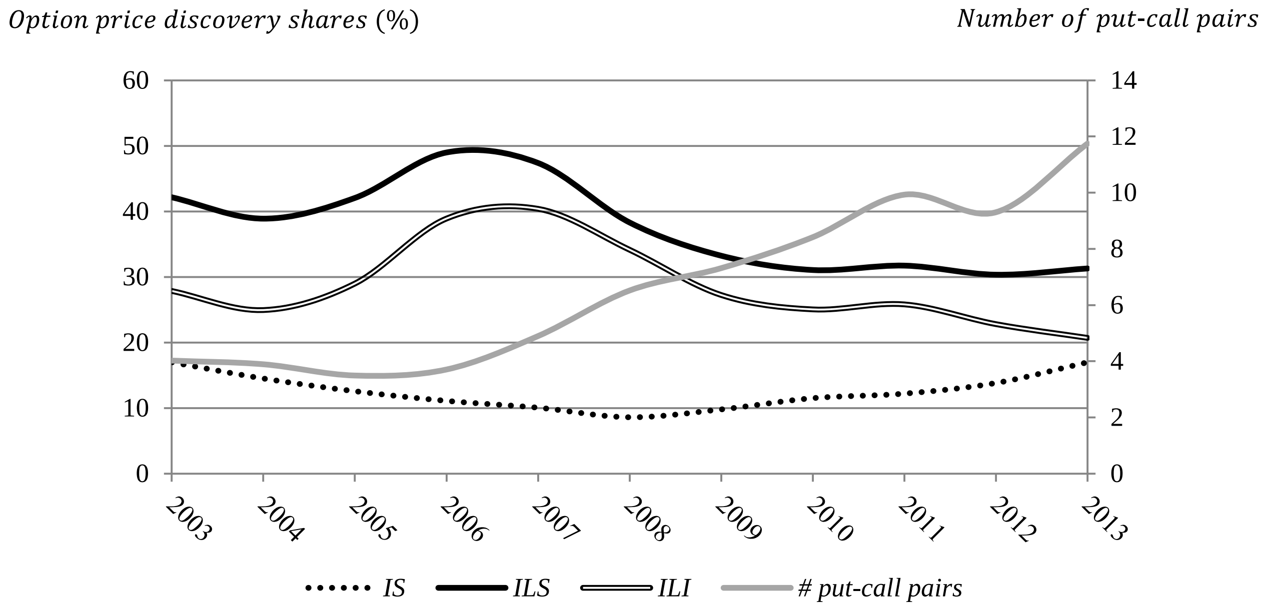

Table 1. Descriptive statistics

This table reports the descriptive statistics for our sample of 35 stocks and their options between April 17, 2003 and April 17, 2013. $Spread$ is the daily time-weighted average quoted bid-ask spread in dollars. $Volume$ for options is the daily traded dollar volume multiplied by the options omega to make it comparable to stock dollar volume and summed across valid put-call pairs each stock-day (reported in units of $1,000). $Volume$ for stocks is the daily traded dollar volume per stock (in units of $1,000). $Quotes$ is the daily number of changes to the NBBO quotes (for options, quote changes are summed across valid put-call pairs each stock-day). $Price$ is the daily time-weighted average midquote price. $Volatility$ is the daily standard deviation of one-minute stock midquote returns in basis points. $Leverage$ is the option omega, averaged across valid put-call pairs.

Panel A: Sample options					
	$Spread$	$Volume$	$Quotes$	$Price$	$Leverage$
Mean	0.07	10,397	1,345	1.68	15.90
Median	0.06	2,546	1,087	1.29	13.92
Std. Dev.	0.05	69,857	1,079	1.39	8.64
Panel B: Sample stocks					
	$Spread$	$Volume$	$Quotes$	$Price$	$Volatility$
Mean	0.01	570,401	5,276	40.52	7.78
Median	0.01	378,922	2,781	30.79	6.38
Std. Dev.	0.01	543,890	7,219	34.40	4.76

Table 2. Monte Carlo simulations of options price discovery shares

This table reports means of options information shares estimated on simulated data. Panels A, B, and C report results for (i) Hasbrouck information shares (IS), (ii) Yan-Zivot-Putniņš information leadership shares (ILS), and (iii) information leadership indicator (ILI), respectively. All price discovery measures are expressed as percentages (range 0–100%) and rounded to the nearest percent. The simulated data for stock and options prices are generated using the model described in Appendix B. Rows correspond to the probability (q) that in a given sample (representing a day), options lead stocks (e.g., for the row $q = 0.7$, options lead stocks approximately 70% of the time). Columns correspond to different values of the noise (σ_{s_2}) in options prices (the noise in stock prices is held constant). For every parameter combination (every cell in the table), we simulate 5,000 samples (each sample representing a day) of 21,600 time-series observations (representing intraday one-second frequency observations), estimate the options price discovery shares on each sample, and compute the mean across the 5,000 samples. Bold is used to indicate price discovery share estimates that are greater than 50%.

Panel A: *IS* metric			σ_{s_2}			
q	10	13	16	19	22	25
0	02	01	01	01	01	00
0.1	05	03	02	02	01	01
0.2	09	06	04	03	02	02
0.3	12	08	06	04	03	03
0.4	15	10	07	05	04	03
0.5	19	13	09	07	05	04
0.6	22	15	10	08	06	05
0.7	26	17	12	09	07	05
0.8	29	19	14	10	08	06
0.9	32	22	15	11	09	07
1	36	24	17	12	09	07
Panel B: *ILS* metric			σ_{s_2}			
q	10	13	16	19	22	25
0	13	14	14	15	15	16
0.1	20	20	21	21	22	22
0.2	27	27	27	27	28	28
0.3	33	33	34	34	34	35
0.4	40	40	40	40	40	41
0.5	47	47	46	47	47	47
0.6	**53**	**53**	**53**	**53**	**53**	**53**
0.7	**60**	**60**	**59**	**59**	**59**	**59**
0.8	**67**	**66**	**66**	**66**	**65**	**65**
0.9	**74**	**73**	**72**	**72**	**72**	**72**
1	**80**	**79**	**79**	**78**	**78**	**78**
Panel C: *ILI* metric			σ_{s_2}			
q	10	13	16	19	22	25
0	00	00	00	00	00	01
0.1	10	10	10	10	10	10
0.2	20	20	20	20	20	21
0.3	30	30	30	30	30	30
0.4	40	40	40	40	40	40
0.5	**50**	**50**	**50**	**50**	**50**	**50**
0.6	**60**	**60**	**60**	**60**	**60**	**60**
0.7	**70**	**70**	**70**	**70**	**70**	**70**
0.8	**80**	**80**	**80**	**80**	**80**	**80**
0.9	**90**	**90**	**90**	**90**	**90**	**90**
1	**100**	**100**	**100**	**100**	**100**	**100**

Table 3. The effect of noise on options price discovery measures

This table reports estimates of how options price discovery measures are impacted by noise. The options price discovery measures are the information leadership indicator (ILI_{it}) and the Hasbrouck information share (IS_{it}). We use several measures of the noise in options prices. $OptionNoise_{it}$ ($StockNoise_{it}$) is the mean absolute difference between the options-implied stock price (stock price) and the estimated common efficient price. $TotalNoise_{it}$ is the sum of $OptionNoise_{it}$ and $StockNoise_{it}$. $OptionNoiseRatio_{it}$ ($StockNoiseRatio_{it}$) is the ratio of $OptionNoise_{it}$ ($StockNoise_{it}$) to $TotalNoise_{it}$. $OptionSpread_{it}$ is the ratio of the time-weighted average quoted options spread to the sum of the options and stock bid-ask spreads. $OptionTickSize_{it}$ is the ratio of the options tick size to the sum of the options and stock tick sizes. Panel A reports the means of $OptionNoiseRatio_{it}$ and $StockNoiseRatio_{it}$ for deciles of the ($ILI_{it} - IS_{it}$) differential. Panel B reports the coefficient estimates from the following regression of stock-day observations:

$$(ILI_{it} - IS_{it}) = \alpha_i + \beta\, NoiseMeasure_{it} + \sum_j \gamma_j\, Controls_{j,it} + \varepsilon_{it}.$$

$NoiseMeasure_{it}$ is $OptionNoiseRatio_{it}$ in Model 1, $OptionSpread_{it}$ in Model 2, and $OptionTickSize_{it}$ in Model 3. ${DV_t}^{GF}$ and ${DV_t}^{OMM}$ are dummy variables equal to one following the removal of the grandfathering provision and removal of the option market-maker exemption, respectively. $Trend_t$ is a linear time trend. The sample comprises 35 stocks during the April 17, 2003 to April 17, 2013 period. Standard errors are clustered both by stock and date and t-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Decile sort by ($\boldsymbol{ILI_{it} - IS_{it}}$)

$Decile$	N	$OptionNoiseRatio_{it}$	$StockNoiseRatio_{it}$
1	5,471	0.52	0.48
2	5,471	0.67	0.33
3	5,472	0.72	0.28
4	5,471	0.76	0.24
5	5,472	0.80	0.20
6	5,471	0.84	0.16
7	5,472	0.90	0.10
8	5,471	0.92	0.08
9	5,472	1.00	0.00
10	5,471	1.00	0.00

Panel B: Determinants of ($\boldsymbol{ILI_{it} - IS_{it}}$)

	Model 1	Model 2	Model 3
$Intercept$	-156.50	-77.70	-53.52
	(-24.97)***	(-6.74)***	(-4.46)***
$NoiseMeasure_{it}$	184.40	69.32	38.30
	(48.58)***	(8.16)***	(6.02)***
${DV_t}^{GF}$	3.80	4.87	5.98
	(3.74)***	(2.78)***	(3.20)***
${DV_t}^{OMM}$	-2.15	-12.67	-13.62
	(-2.95)***	(-7.66)***	(-7.43)***
$Trend_t$	3.06	6.26	6.80
	(4.38)***	(5.05)***	(4.84)***
R^2 (%)	49.79	5.22	4.60
Fixed Effects	Stock	Stock	Stock

Table 4. Options market price discovery shares

This table reports options market price discovery shares for our sample of 35 stocks between April 17, 2003 and April 17, 2013. The price discovery shares are: (i) Hasbrouck information share (IS), (ii) Yan-Zivot-Putniņš information leadership share (ILS) and (iii) information leadership indicator (ILI). All price discovery measures are expressed as percentages (range 0–100%). UML is the difference between the upper and lower bound estimates for IS. N is the number of stock-day observations and differs across tickers because (i) when a stock changes ticker, is acquired or is de-listed, it is removed from our sample, and (ii) not all stock-days have valid put-call pairs that satisfy the sampling criteria in Section 2. CI is a bootstrapped 95% confidence interval. σ_{Total} is the standard deviation of the price discovery estimates across all stock-days. σ_{ε} is the standard deviation of measurement error. σ_{PD} is the lower bound of the standard deviation of true variation (not measurement error) in the price discovery measure across stock-days.

Panel A: Pooled sample descriptive statistics

	N	IS	ILS	ILI	UML
Mean	54,714	11.84	36.75	28.94	1.74
CI	54,714	(3.99, 39.43)	(26.82, 56.29)	(25.90, 44.58)	
Median	54,714	8.01	14.04	0.00	1.08

Panel B: Decomposition of the variance in price discovery estimates

	N	IS	ILS	ILI
σ_{Total}	54,714	12.50	40.56	45.17
σ_{ε}	54,714	8.86	7.37	4.67
σ_{PD}	54,714	8.81	39.89	44.93

Panel C: Pooled sample descriptive statistics by ticker

$Ticker$	N	IS	ILS	ILI	$Ticker$	N	IS	ILS	ILI
AIG	1,911	12.16	38.33	30.35	HD	1,923	10.70	32.77	24.17
AMAT	1,299	10.20	34.60	27.68	IBM	1,907	10.91	37.18	28.90
AMGN	1,625	12.60	41.58	36.00	INTC	1,203	11.54	21.44	13.88
AMR	1,928	14.65	40.10	29.02	JPM	1,880	10.15	31.05	23.56
AMZN	1,586	7.94	35.70	31.15	KLAC	1,651	8.97	52.18	47.12
AOL	90	18.41	43.19	27.22	MMM	2,222	11.37	47.74	41.13
BMY	1,819	11.30	31.25	22.62	MO	2,063	13.58	48.55	40.45
BRCM	1,493	7.96	40.19	35.43	MSFT	1,409	11.20	40.05	34.39
C	1,701	12.08	33.26	25.22	MWD	475	9.15	36.19	23.37
COF	2,403	10.61	42.36	34.33	ORCL	1,449	10.67	31.72	24.29
CPN	388	18.00	50.16	42.65	PFE	1,445	13.40	29.24	20.80
CSCO	1,331	11.96	26.60	18.63	QCOM	1,437	8.73	31.66	26.55
DELL	1,424	11.52	28.10	20.86	QLGC	1,915	14.26	42.12	32.98
EBAY	1,565	9.25	33.31	27.80	SBC	307	9.17	36.40	27.20
EMC	2,342	15.50	32.10	20.64	TYC	2,054	14.93	40.72	31.99
F	1,984	15.77	30.60	20.56	XLNX	1,337	11.54	38.95	32.87
GE	1,841	12.65	31.94	22.98	XOM	1,658	10.82	35.88	29.07
GM	1,649	12.78	39.75	30.47					

Table 5. Options market price discovery around SEC insider trading prosecutions

This table reports coefficient estimates from the following regression using stock-day observations:

$$PD_{it} = \alpha_i + \beta_1 OptionDV_{it} + \beta_2 StockDV_{it} + \beta_3 BothDV_{it} + \sum_j \gamma_j\, Controls_{j,it} + \varepsilon_{it}.$$

PD_{it} is the price discovery measure for the options market using the Hasbrouck information share (IS_{it}) in Model 1, Yan-Zivot-Putniņš information leadership share (ILS_{it}) in Model 2, and information leadership indicator (ILI_{it}) in Model 3. $OptionDV_{it}$ is a dummy variable equal to one if illegal insider trading occurs in the options market only, $StockDV_{it}$ is a dummy variable equal to one if illegal insider trading occurs in the stock market only, and $BothDV_{it}$ is a dummy variable equal to one if illegal insider trading occurs in both the stock and options markets. $Trend_t$ is a linear time trend. The sample comprises 36 stocks in which insiders were successfully prosecuted for illegal trading between January 1, 1999 and August 30, 2014. For each announcement we take one month before the first insider trade, one month after the last insider trade, and all stock-days in between to give a sample that includes days with and without illegal insider trades. Standard errors are clustered both by stock and date and t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels.

	IS_{it} Model 1	ILS_{it} Model 2	ILI_{it} Model 3
$Intercept$	27.51	78.98	35.33
	(0.58)	(1.24)	(0.92)
$OptionDV_{it}$	-3.93	7.12	6.14
	(-1.60)	(2.77)***	(2.14)**
$StockDV_{it}$	-2.20	-0.56	-1.34
	(-0.91)	(-0.15)	(-0.33)
$BothDV_{it}$	5.77	7.34	6.41
	(1.11)	(0.95)	(0.64)
$Trend_t$	0.59	-5.71	-1.80
	(0.09)	(-0.68)	(-0.36)
R^2 (%)	16.04	8.20	6.30
Fixed effects	Stock	Stock	Stock

Table 6. Options market price discovery around information events

This table reports options price discovery shares around information events. The price discovery shares are: (i) Hasbrouck information share (IS), (ii) Yan-Zivot-Putniņš information leadership share (ILS), and (iii) information leadership indicator (ILI). All price discovery measures are expressed as percentages (range 0–100%). N is the number of stock-day observations. $Diff$ is the difference in mean options price discovery during information events (insider trading stock-days in Panel A and price-sensitive news in Panel B) and other non-information times (stock-days that do not have the corresponding type of information event). The t-statistic is the Satterthwaite t-statistic for the difference in means. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. In Panel A, the information events are the 212 stock-days that are known to have illegal insider trading, which occur in 36 stocks between January 1, 1999 and August 30, 2014. In Panel B, the information events are the 3,930 stock-days in our sample for which there is a price-sensitive news release.

Panel A: Options price discovery during illegal insider trading				
	N	IS	ILS	ILI
Mean	212	25.09	50.93	41.04
Difference	54,714	13.25	14.18	12.10
t-statistic	54,714	7.91***	5.06***	3.57***
Panel B: Options price discovery during price-sensitive news releases				
	N	IS	ILS	ILI
Mean	3,930	11.32	38.92	32.04
Difference	50,784	-0.56	2.33	3.33
t-statistic	50,784	-2.82**	3.35***	4.33***

Table 7. Options market price discovery during price-sensitive news releases

This table reports coefficient estimates from the following regression using stock-day observations:

$$PD_{it} = \alpha_i + \beta News_{it} + \sum_j \gamma_j Controls_{j,it} + \varepsilon_{it}.$$

PD_{it} is the price discovery measure for the options market using the Hasbrouck information share (IS_{it}) in Model 1, Yan-Zivot-Putniņš information leadership share (ILS_{it}) in Model 2, and information leadership indicator (ILI_{it}) in Model 3. $News_{it}$ is a dummy variable equal to one when price-sensitive news relating to the stock is released. A stock-day has a price-sensitive news release if: (i) the number of news items on the given day mentioning the given stock exceed the median number of news releases per stock-year during our sample period, and (ii) the average relevance score for the news on a given stock-day exceeds 0.75. DV_t^{GF} and DV_t^{OMM} are dummy variables equal to one following the removal of the Grandfathering provision and removal of the option market-maker exemption, respectively. $Trend_t$ is a linear time trend. The sample comprises 35 stocks during the period April 17, 2003 to April 17, 2013. Standard errors are clustered both by stock and date and t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels.

	IS_{it} Model 1	ILS_{it} Model 2	ILI_{it} Model 3
$Intercept$	18.82	29.38	2.91
	(7.64)***	(5.13)***	(0.35)
$News_{it}$	-0.58	3.50	4.19
	(-2.49)**	(5.13)***	(5.51)***
DV_t^{GF}	-2.27	-5.29	0.15
	(-4.87)***	(-4.03)***	(0.09)
DV_t^{OMM}	3.84	-10.28	-13.64
	(8.19)***	(-7.65)***	(-9.01)***
$Trend_t$	-1.23	2.24	4.77
	(-3.38)***	(2.56)**	(3.76)***
R^2 (%)	3.92	4.65	3.69
Fixed Effects	Stock	Stock	Stock

Table 8. Determinants of price discovery

This table reports coefficient estimates from the following regression using stock-day observations:

$$PD_{it} = \alpha_i + \beta_1 Spread_{it} + \beta_2 Volume_{it} + \beta_3 Volatility_{it} + \sum_j \gamma_j\, OtherDeterminants_{j,it} + \varepsilon_{it}.$$

PD_{it} is the price discovery measure for the options market using the Hasbrouck information share (IS_{it}) in Model 1, Yan-Zivot-Putniņš information leadership share (ILS_{it}) in Model 2, and information leadership indicator (ILI_{it}) in Models 3 to 6. $Spread_{it}$ is the ratio of the time-weighted average quoted bid-ask spread in the options market to that of the stock market. $Volume_{it}$ is the ratio of options omega-adjusted dollar volume to stock dollar volume. $Volatility_{it}$ is the standard deviation of one-minute stock midquote returns. $Quotes_{it}$ is the ratio of the number of NBBO quote changes in the options market to that of the stock market. $Leverage_{it}$ is the omega of the call option (if the next day's stock return is positive) or the omega of the put option (if the next day's stock return is negative). $StrikeDistance_{it}$ is the absolute difference between the underlying stock price and the options strike price. $ImpVol_{it}$ is the implied volatility from the Black-Scholes model. DV_{it}^{a-b} is a dummy variable equal to one if the time to expiry is between a and b days, inclusive (options with expiry between 40 and 70 days are the base case). ${DV_t}^{GF}$ and ${DV_t}^{OMM}$ are dummy variables equal to one following the removal of the Grandfathering provision and removal of the option market-maker exemption, respectively. $Trend_t$ is a linear time trend. All explanatory variables are in natural log form (except dummy variables). The sample comprises 35 stocks during the period April 17, 2003 to April 17, 2013. Standard errors are clustered both by stock and date and t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels.

	IS_{it} Model 1	ILS_{it} Model 2	ILI_{it} Model 3	ILI_{it} Model 4	ILI_{it} Model 5	ILI_{it} Model 6
$Intercept$	-17.57	18.60	41.47	30.66	37.29	14.43
	(-4.05)***	(2.16)**	(3.99)***	(2.99)***	(4.58)***	(1.13)
$Spread_{it}$	-0.09	13.04	11.74	7.22	7.45	7.83
	(-0.16)	(10.17)***	(7.33)***	(5.28)***	(5.64)***	(5.80)***
$Volume_{it}$	-0.59	1.09	1.34	0.78	0.55	0.32
	(-3.95)***	(3.15)***	(3.37)***	(1.99)**	(1.42)	(0.75)
$Volatility_{it}$	-3.30	-1.32	2.31	2.26	3.55	4.36
	(-6.06)***	(-1.18)	(1.73)*	(1.80)*	(2.96)***	(4.19)***
$Quotes_{it}$				-4.33	-3.72	-2.52
				(-5.01)***	(-4.22)***	(-3.09)***
$Leverage_{it}$				2.59		2.64
				(3.10)***		(3.47)***
$StrikeDistance_{it}$					0.76	
					(1.39)	
$ImpVol_{it}$					-4.11	
					(-1.64)*	
DV_{it}^{0-9}					6.10	
					(3.34)***	
DV_{it}^{10-19}					4.13	
					(1.93)*	
DV_{it}^{20-29}					3.14	
					(1.81)*	
DV_{it}^{30-39}					1.98	
					(1.80)*	
${DV_t}^{GF}$						0.73
						(0.26)
${DV_t}^{OMM}$						-7.30
						(-3.89)***
$Trend_t$						4.41
						(3.36)***
R^2 (%)	2.13	4.06	2.83	3.43	3.50	5.41
Fixed Effects	None	None	None	None	None	Stock

Table 9. Two-stage least-squares IV regressions estimating the impact of liquidity on price discovery

This table reports coefficient estimates from the first and second stages of an instrumental variables model. The first stage models (Models 1 and 2) use the reduction in options tick sizes (due to the Penny Pilot Program) as an instrument for the relative level of liquidity in options ($Spread_{it}$, the ratio of the time-weighted average quoted bid-ask spread in the options market to that of the stock market). $DV_{it}{}^{Tick}$ is a dummy variable equal to one after the reduction in the options tick size. $Volume_{it}$ is the ratio of options omega-adjusted dollar volume to the stock dollar volume. $Volatility_{it}$ is the standard deviation of one-minute stock midquote returns. $Quotes_{it}$ is the ratio of the number of NBBO quote changes in the options market to that of the stock market. $Leverage_{it}$ is the omega of the call option (if the next day's stock return is positive) or the omega of the put option (if the next day's stock return is negative). $DV_t{}^{GF}$ and $DV_t{}^{OMM}$ are dummy variables equal to one following the removal of the grandfathering provision and removal of the option market-maker exemption, respectively. $Trend_t$ is a linear time trend. The second stage models (Models 3 and 4) regress ILI_{it} (the options market information leadership indicator for stock i on day t) on the fitted value of options market liquidity ($\widehat{Spread_{it}}$) and control variables. All variables are in natural log form (except dummy variables). The sample comprises 35 stocks during the period April 17, 2003 to April 17, 2013. Standard errors are clustered both by stock and date and t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels.

	$Spread_{it}$ Model 1	$Spread_{it}$ Model 2	ILI_{it} Model 3	ILI_{it} Model 4
$Intercept$	53.70	-63.66	30.14	15.52
	(1.88)*	(-2.28)**	(2.83)***	(1.18)
$DV_{it}{}^{Tick}$	-62.14	-57.03		
	(-7.58)***	(-6.40)***		
$\widehat{Spread_{it}}$			3.22	4.81
			(1.03)	(1.44)
$Volume_{it}$	-3.79	-3.76	0.29	0.21
	(-7.23)***	(-7.54)***	(0.67)	(0.49)
$Volatility_{it}$	-13.41	-13.16	2.51	3.81
	(-4.02)***	(-3.82)***	(1.82)*	(3.25)***
$Quotes_{it}$	-22.85	-23.51	-5.27	-3.60
	(-10.11)***	(-8.68)***	(-3.80)***	(-2.61)***
$Leverage_{it}$	-14.05	-14.28	2.23	2.11
	(-4.44)**	(-4.58)**	(3.22)***	(2.50)**
DV_{it}^{GF}		-4.14		-0.20
		(-0.96)**		(-0.08)
DV_{it}^{OMM}		-0.11		-7.39
		(-0.03)		(-3.88)***
$Trend_t$		-1.21		4.21
		(-0.58)		(3.11)***
R^2 (%)	78.03	78.08	4.80	5.10
Fixed Effects	Stock	Stock	Stock	Stock

Table 10. Relation between options price discovery shares and price adjustments following disagreement events

This table reports coefficient estimates from the following regression using stock-day observations:

$$PD_{it} = \alpha_i + \beta_1 Adjust_{it} + \sum_j \gamma_j \, OtherDeterminants_{j,it} + \varepsilon_{it}.$$

PD_{it} is the price discovery measure for the options market using the Hasbrouck information share (IS_{it}) in Model 1, the Yan-Zivot-Putniņš information leadership share (ILS_{it}) in Model 2, and the information leadership indicator (ILI_{it}) in Models 3. $Adjust_{it}$ is the ratio of adjustment in stock prices to the adjustment in options prices in the ten seconds following price disagreement events. $Spread_{it}$ is the ratio of the time-weighted average quoted bid-ask spread in the options market to that of the stock market. $Volume_{it}$ is the ratio of options omega-adjusted dollar volume to stock dollar volume. $Volatility_{it}$ is the standard deviation of one-minute stock midquote returns. $Quotes_{it}$ is the ratio of the number of NBBO quote changes in the options market to that of the stock market. $Leverage_{it}$ is the omega of the call option (if the next day's stock return is positive) or the omega of the put option (if the next day's stock return is negative). ${DV_t}^{GF}$ and ${DV_t}^{OMM}$ are dummy variables equal to one following the removal of the Grandfathering provision and removal of the option market-maker exemption, respectively. $Trend_t$ is a linear time trend. All explanatory variables are in natural log form (except dummy variables). The sample comprises 35 stocks during the period April 17, 2003 to April 17, 2013. Standard errors are clustered both by stock and date and t-statistics are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels.

	IS_{it} Model 1	ILS_{it} Model 2	ILI_{it} Model 3
$Intercept$	-6.29	22.88	15.10
	(-1.08)	(2.53)**	(1.21)
$Adjust_{it}$	-0.40	2.82	3.65
	(-0.65)	(2.26)**	(2.54)**
$Spread_{it}$	1.82	9.77	7.77
	(1.55)	(8.13)***	(5.65)***
$Volume_{it}$	-0.20	0.32	0.31
	(-1.87)*	(0.93)	(0.71)
$Volatility_{it}$	-3.20	1.62	4.46
	(-4.93)***	(1.84)*	(4.40)***
$Quotes_{it}$	2.19	0.38	-2.38
	(5.46)***	(0.51)	(-2.95)***
$Leverage_{it}$	-0.20	1.39	2.59
	(-0.61)	(1.83)*	(3.40)***
DV_{it}^{GF}	1.24	-2.61	0.49
	(1.81)*	(-1.17)	(0.17)
DV_{it}^{OMM}	1.07	-6.87	-7.37
	(1.92)*	(-3.97)***	(-3.92)***
$Trend_t$	-1.24	2.29	4.46
	(-3.54)***	(2.45)**	(3.43)***
R^2 (%)	6.73	6.11	5.46
Fixed Effects	Stock	Stock	Stock