Author's Accepted Manuscript

Dark trading and price discovery

Carole Comerton-Forde, Tālis J. Putniņš

Cite this article as: Carole Comerton-Forde, Tālis J. Putniņš, Dark trading and price discovery, *Journal of Financial Economics*, http://dx.doi.org/10.1016/j.jfineco.2015.06.013

Dark trading and price discovery[*]

Carole Comerton-Forde [a,] and Tālis J. Putniņš [b,c]

[a] *Department of Finance, University of Melbourne, 198 Berkeley Street, Parkville, VIC 3010, Australia.*
[b] *University of Technology Sydney, PO Box 123 Broadway, NSW 2007, Australia.*
[c] *Stockholm School of Economics in Riga, Strelnieku Street 4a, Riga, LV 1010, Latvia.*

February 13, 2015

Abstract

Regulators globally are concerned that dark trading harms price discovery. We show that dark trades are less informed than lit trades. High levels of dark trading increase adverse selection risk on the lit exchange by increasing the concentration of informed traders. Using both high- and low-frequency measures of informational efficiency we find that low levels of non-block dark trading are benign or even beneficial for informational efficiency, but high levels are harmful. In contrast, we find no evidence that block trades in the dark impede price discovery.

JEL classification: G14

Keywords: dark pool, price discovery, efficiency

1. Introduction

Technology has transformed trading and new trading venues, known as dark pools, have emerged. Although the term "dark trading" is new, the concept is not. Dark trading of old includes block orders managed by upstairs brokers and orders in the pockets of floor brokers not yet revealed to the market. Today's dark pools systematically match orders without providing any pre-trade transparency. Technology has facilitated rapid growth in dark trading around the world. For example, dark trading's share of US consolidated volume has grown from 17% in July 2008 to 37% in June 2014 (Rosenblatt Securities). Technology has also changed the nature of dark trading, with block executions becoming less significant than non-block dark executions, due to market participants' increasing use of algorithms to execute dark trades.[1]

Many regulators and stock exchanges have expressed concern that excessive growth in dark trading can harm price discovery. For example, in a recent speech, Securities and Exchange Commission (SEC) Chairman White states "…we must continue to examine whether dark trading volume is approaching a level that risks seriously undermining the quality of price discovery provided by lit venues." Over the last five years, many regulators have undertaken public consultations or proposed new regulations on dark trading. However, to date, only the Canadian and Australian regulators have implemented new rules.[2] The extensive consultation and subsequent lack of action by regulators reflects the uncertainty about the real costs and benefits of dark trading and the competing interests of the different participants in the market. In many cases this uncertainty is compounded by an inability to accurately identify and measure dark trading, making it difficult to assess its impact. This has also limited academic research on dark trading.

This paper is the first to focus empirically on the effect of dark/block trading on price discovery. This focus is consistent with the intense global regulatory concern about

---

[1] Seppi (1990) reports that block trades accounted for roughly half of the NYSE trading volume in 1989, compared with current NYSE statistics which report that blocks account for 25% of volume. Tuttle (2013) shows that during May 2012, the distribution of trade sizes executed in dark Alternative Trading Systems is very similar to that of exchanges.

[2] For example, in November 2009, the US SEC proposed rules on the "Regulation of Non-Public Trading Interest" but to date, no rule changes have been made in the US. In Europe, Markets in Financial Instruments Directive II (MiFID II) proposes the introduction of a double cap on dark trading with a 4% cap on trading in a single dark venue and an 8% cap on total dark trading across all venues. The Canadian and Australian securities regulators imposed price improvement requirements for non-block dark trades on October 15, 2012 and May 27, 2013, respectively.

the impact of dark/block trading on price discovery. Our analysis overcomes the issues in accurately observing and measuring dark trading in US and other markets by using highly granular equities data from the Australian Securities Exchange (ASX). Our data enable us to precisely identify and measure all dark and block trading over a long time-series for a broad cross-section of stocks. All orders and trades are time-stamped to the millisecond and the time-stamps are consistent across the different trading mechanisms. The data also enable us to distinguish between different types of trades, allowing us to assess differences between the traditional 'upstairs' block trades and smaller dark trades executed without negotiation. This distinction is important because in most markets regulators apply different rules to the two types of dark trading, which are thought to have different impacts. For the remainder of the paper we refer to non-block dark trading as 'dark' trading, and block dark trading as 'block' trading.

We address three questions. First, where are informed and uninformed trades typically executed and thus, how informative are lit, dark, and block trades? Second, how does the level of dark and block trading impact adverse selection risk on the lit exchange? Third, what is the association between dark/block trading and price discovery? Given that regulators are typically only concerned about high levels of dark/block trading, we examine whether the association between dark/block trading and price discovery is nonlinear and whether there is a threshold or 'tipping point' above which dark/block trading is harmful.[3]

Our empirical work is guided by two recent theoretical models of how dark trading impacts price discovery (Ye, 2012; Zhu, 2014). These models reach conflicting conclusions, with Ye predicting a negative association, and Zhu a positive association between price discovery and dark trading. Using theory to predict the effects of dark trading is difficult because it concurrently affects: (i) the degree of transparency; (ii) the extent of segmentation of informed and uninformed traders; and (iii) fragmentation of trading, which changes the way traders submit orders. Each of these three characteristics can be benign, beneficial, or detrimental to price discovery, depending on the circumstances. Our empirical analysis helps resolve the conflicting theory and understand which of the possible mechanisms dominates.

---

[3] For example, in testimony before the Senate Banking, Housing and Urban Affairs Committee, Nasdaq OMX Chief Economist, Frank Hatheway, states that based on Nasdaq's empirical analysis "execution quality begins to deteriorate when stocks experience dark trading in excess of 40% of total volume."

Our results support the hypothesis that dark trading leads to partial segmentation of informed and uninformed traders, as predicted by Zhu (2014). We find that orders executed in the dark tend to be less informed than orders executed in the lit market, consistent with informed traders facing lower execution probabilities in the dark than uninformed traders. By disproportionately reducing the number of uninformed trades in the lit market, high levels of dark trading increase adverse selection risk in the lit market, leading to wider bid-ask spreads, consistent with Zhu (2014). The reduction in uninformed traders in the lit market, accompanied by wider spreads, reduces incentives for costly information acquisition given that informed traders are less able to trade in the dark than uninformed traders. Therefore, dark trading could decrease the aggregate amount of information produced about fundamental values.

We also find that as dark trading increases, order book quotes take on a more important role in impounding new information compared to trade prices, consistent with liquidity providers in the lit market becoming increasingly informed. This result is consistent with the notion that high levels of dark trading increase adverse selection risks in the lit market and informed traders have a comparative advantage in providing liquidity when adverse selection risks are high, due to their informational advantage. We also find that dark trades play a greater role in price discovery as the level of dark trading increases. However, the dark market's share of price discovery increases at a slower rate than its increase in market share, providing further evidence that dark trades tend to be less informed than lit trades.

Finally, our results show that dark and block trades have different impacts on informational efficiency. Low levels of dark trading are either benign or beneficial, but high levels are harmful to informational efficiency. The deterioration in informational efficiency begins to occur when dark trading in a given stock exceeds approximately 10% of dollar volume. The change in informational efficiency when dark trading exceeds this 'tipping point' is economically meaningful. For a typical stock, the level of dark trading is below harmful levels. In contrast, we find no evidence that block trades negotiated away from the exchange without pre-trade transparency harm informational efficiency. In fact, having some block trades execute away from the lit market (up to approximately 40% of dollar volume) can be beneficial to informational efficiency. The benefits could

4

be due to upstairs block brokers tapping into liquidity that would not otherwise be expressed in the limit order book.

We address the endogeneity of dark/block trading using instrumental variables, and, therefore, provide evidence on the causal relation between dark/block trading and informational efficiency. Our results are robust to a number of control variables, hold in both large and small stocks, and in early and later parts of our sample. Our analysis is robust to the use of high-frequency informational efficiency metrics calculated from intraday observations as well as lower frequency metrics computed using daily observations.

This paper is related to recent empirical studies of dark trading, most of which analyze the relation between dark trading and liquidity. For example, Buti, Rindi, and Werner (2011) use data from 11 out of 32 US dark pools and conclude that dark pools improve spreads, depth, and short-term volatility. Ready (2013) examines the determinants of volume in two block dark pools. He finds that these pools execute a lower fraction of institutional volume in stocks with higher levels of adverse selection. Nimalendran and Ray (2014) examine data from one of the 32 US dark pools and find that trading in the dark pool is associated with increased spreads and price impact on the quoting exchanges. Kwan, Masulis, and McInish (2015) use a comprehensive data set of US off-exchange trading classified into five types of dark venues to show that the un-level playing field between dark venues and exchanges increases fragmentation and has a detrimental impact on liquidity.

Our paper extends the recent empirical literature on dark trading and liquidity with an analysis of dark trading and price discovery. While both liquidity and price discovery are fundamentally important functions of the market, theory suggests that the impact of dark trading on price discovery can be entirely different to the effects on liquidity e.g., Zhu (2014) predicts opposite effects on liquidity and price discovery.

2. Theory and previous literature

The impact of dark and block trading on price discovery is a complex issue because they simultaneously affect: (i) the level of transparency; (ii) fragmentation of order flow across multiple trading venues; and (iii) segmentation of informed and uninformed order flow. This section reviews the theory on each of these characteristics,

focusing on the effects on price discovery. Dark and block trading are considered separately because the literature suggests that they have different impacts and because market regulators tend to view these two forms of non-transparent trading differently.

2.1. Dark trading

Dark and lit trades differ in their levels of pre-trade transparency, but not post-trade transparency.[4] For example, a limit order submitted to a lit exchange is immediately visible to all market participants and thus has an immediate price impact as market participants revise their beliefs about the fundamental value.[5] In contrast, if the limit order instead rests in a dark market, no one except the order submitter can observe the order and none of the information contained in the limit order can be impounded into prices until a trade occurs. If the limit order does not execute, the market will never know about the order. Even if the order eventually executes and market participants observe the dark trade printed to the tape, the market still knows less about the order than if the order had been sent to the lit market. For example, the market does not know the time at which the limit order was submitted, the original size of the limit order, any revisions to the order price, and the venue in which it was executed, all of which can be informative. Furthermore, market participants can usually determine the direction of trades in a lit market (the trade initiator) because trades generally execute at the best bid or ask price. In contrast, trades in the dark can occur within the spread, making it difficult for the market to infer the trade direction. As all of these examples illustrate, market participants observe less information about order flow sent to dark venues than order flow sent to the lit market. Because order flow conveys information e.g., in Roşu (2013) and Kaniel and Liu (2006), both market and limit orders convey information, dark trading could have a negative impact on price discovery.

The conjecture that a reduction in transparency can harm price discovery is supported by the literature. Most, but not all, previous studies argue that transparency benefits liquidity and price discovery. For example, in various auction and dealer markets examined by Pagano and Röell (1996), pre-trade transparency allows market

---

[4] Pre-trade transparency is the degree to which information is available to market participants about buying or selling interest, including quotes to buy or sell and the volumes available at the quotes. Post-trade transparency is the degree and timeliness of information about trades after they execute.

[5] Hautsch and Huang (2012), among others, find empirical evidence that limit orders have price impacts.

6

makers to learn information from trades more quickly, leading to more efficient prices and lower trading costs for uninformed traders. In Baruch's (2005) model, making the limit order book transparent increases the ability of market participants to compete with the specialist in liquidity provision. The increased competition increases liquidity and improves price discovery. In contrast, Boulatov and George (2013) find that hiding liquidity-providing orders causes more aggressive competition among informed traders in providing liquidity, which improves price discovery. The empirical evidence, although not unanimous, also tends to support the view that pre-trade transparency in most circumstances has positive effects on price discovery.[6]

Eom et al. (2007) argue that market quality is an increasing concave function of pre-trade transparency, or equivalently, a *decreasing* concave function of pre-trade *opaqueness*. Dark trading increases pre-trade opaqueness in a market and therefore high levels of dark trading could harm price discovery. If market quality is a decreasing concave function of pre-trade opaqueness, low levels of dark trading should not be harmful to price discovery and could even be beneficial.

In addition to reduced transparency, the launch of new, automated dark venues fragments order flow and trading activity. Market participants typically use smart order routers to consider multiple venues when trying to get their orders filled and rest orders in multiple venues. Fragmentation has both positive and negative effects on market quality, including price discovery. Network externalities suggest there are benefits to consolidation.[7] When more traders use a particular market, the market's ability to match buyers and sellers increases. Consequently, trading costs decrease, which attracts more traders. Improved liquidity incentivises production of costly information and facilitates arbitrage, and thus can increase the informativeness of prices (e.g., Kyle, 1984; Chordia, Roll, and Subrahmanyam, 2008). Fragmentation can also harm liquidity and price discovery by increasing search costs and thus decreasing competition between liquidity

---

[6] For example, Boehmer, Saar, and Yu (2005) provide empirical support for Baruch's (2005) model examining the increase pre-trade transparency resulting from the introduction of NYSE's OpenBook. However, Madhavan, Porter, and Weaver (2005) report that a similar increase in pre-trade transparency in Toronto decreased liquidity. They argue that too much pre-trade transparency makes traders reluctant to post limit orders because of the increased "free option" cost. Hendershott and Jones (2005) examine trading in three exchange-traded funds (ETFs) and find that when the Island Electronic Communication Network (ECN) suspends the display of the limit order book, overall trading costs increase and price discovery deteriorates.

[7] See, for example, Chowdhry and Nanda (1991), Madhavan (1995), and Hendershott and Mendelson (2000).

providers (e.g., Yin, 2005). Fragmentation can also benefit markets through increased competition between trading venues, which can lower trading costs (e.g., Foucault and Menkveld, 2008; Colliard and Foucault, 2012). The empirical evidence on the effects of fragmentation is mixed, mirroring the various opposing effects predicted by theory.[8]

Segmentation refers to the tendency for different types of traders to use different markets and can impact on price discovery. There are several reasons why dark trading can lead to segmentation of informed and uninformed order flow. First, at any point in time, informed traders are more likely than uninformed traders to cluster on one side of the market (either buying or selling). Consequently, informed traders face lower execution probability in the dark than uninformed traders (Zhu, 2014). Second, in some jurisdictions, including the US and Australia, dark venues are subject to lower regulatory requirements regarding fair access and consequently can discourage or exclude relatively informed order flow (Boni, Brown, and Leach, 2012). Third, dark trading can make it easier for brokers to internalize order flow. Internalization of uninformed order flow is more profitable for a broker than informed order flow due to the differences in adverse selection costs. Therefore, internalization can also lead to a disproportionately large share of uninformed trades taking place in the dark. A disproportionately high proportion of uninformed trades in the dark implies an increased concentration of informed traders in the lit market. It is therefore expected that adverse selection costs and spreads in the lit market increase with the level of dark trading, as predicted by Zhu (2014).

The effects of segmentation on price discovery, however, are less clear. Different theories show that a higher concentration of informed trading can have a positive, zero, or negative effect on price discovery, depending on a number of factors. A substantial decrease in uninformed traders in the lit market could harm price discovery by reducing the profitability of acquiring unique private information (e.g., Kyle, 1981, 1984, 1989). Alternatively, if all informed traders have the same piece of private information as in Zhu (2014), fewer uninformed traders in the lit market could improve price discovery.

---

[8] For example, Hendershott and Jones (2005) find that fragmentation in trading for three ETFs harms liquidity and price discovery, whereas O'Hara and Ye (2011), using off-exchange volume of US stocks as a proxy for fragmentation, find that more fragmented stocks have lower transaction costs and better informational efficiency. Degryse, de Jong, and van Kervel (2014) argue that fragmentation of trading across lit venues can have different impacts than fragmentation stemming from dark venues. Using a sample of Dutch stocks they show that fragmentation across visible order books improves consolidated liquidity, whereas dark trading has a detrimental effect. The different effects of lit and dark fragmentation could also be due to the other factors, including transparency and segregation of order flow.

Therefore, the impact of the level of informed trading in lit venues on price discovery is ultimately an empirical question.

2.2. Block trading

In many markets large block trades are negotiated manually between brokers in the 'upstairs' market. From the perspective of a trader that participates only in the downstairs market, an upstairs block trade has no pre-trade transparency (similar to dark trades). However, from the perspective of the upstairs market trade counterparty, the block trade has greater pre-trade transparency than a lit trade, because in negotiating upstairs trades, brokers are able to signal the likely motivation for the trade, thereby reducing adverse selection risks and execution costs for large liquidity-motivated trades (Madhavan and Cheng, 1997; Bessembinder and Venkataraman, 2004).

Differences in the nature of upstairs block trading and dark trading imply they should have different consequences for price discovery. First, the upstairs market is able to facilitate trades that would not be possible in the downstairs market (Madhavan and Cheng, 1997; Bessembinder and Venkataraman, 2004). Such trades are made possible by the upstairs brokers' ability to tap into unexpressed liquidity of large institutional traders, thereby expanding the total liquidity available to the market (Grossman, 1992), and negotiating prices outside the limit order book quotes (Bessembinder and Venkataraman, 2004). In contrast, if a market for dark trades did not exist, most of the dark trades would simply execute on the lit market. By expanding the total available liquidity and facilitating trades that would not be possible in the limit order book, block trading is likely to benefit price discovery by providing additional information about the fundamental value to market participants (on a post-trade basis). Without the ability to draw out this latent liquidity upstairs, if these orders are sent to the lit market they could create short-term demand and supply imbalances and temporary price distortions (Bessembinder and Venkataraman, 2004). Therefore, block trades are not expected to harm price discovery.

3. Institutional setting

During our sample period, February 2008 to October 2011, the ASX was the only stock exchange operating in Australia. The ASX is one of the top ten equity markets in

the world by market capitalization. There are approximately 2,200 companies listed on the ASX with a market capitalization of around 1.5 trillion Australian dollars (henceforth, '$' and 'dollars' refer to Australian dollars).

The ASX operates a transparent central limit order book (CLOB), in which orders are matched based on price-then-time priority.[9] The CLOB is anonymous, but the brokerage firms associated with each trade are reported to the market on T+3. There are no official market makers. During our sample period, the ASX Operating Rules provided two exceptions that allowed trades to be executed away from the CLOB with reduced pre-trade transparency, provided that the trades are reported to the market immediately. These exceptions include:

(i) *Block trades* that must have a minimum value of $1 million or comprise a portfolio of trades with a combined value of at least $5 million. These trades can be negotiated away from the CLOB at any price, without any requirement to interact with the CLOB.

(ii) *Dark trades* that have no minimum size requirement but must occur at or within the prevailing best bid or ask price. Dark trades comprise trades executed in Australian dark pools, including the ASX-operated dark pool named Centre Point and a large number of broker-operated dark pools, as well as trades manually matched away from the lit market. Centre Point (launched in June 2010) executes orders based on time priority at the midpoint of the bid-ask spread on the CLOB. Centre Point orders do not interact with orders on the CLOB. At the beginning of our sample in February 2008, there were four broker-operated dark pools in Australia. This number grew to 16 by the end of our sample period.

Further details about these exceptions and dark pools in Australia are provided in the Online Appendix.[10]

There were a number of other institutional changes that impacted dark and block trading during our sample period. On November 30, 2009, the ASX Operating Rules

---

[9] Fully hidden orders are not permitted; however, the ASX provides two order types that provide reduced transparency: undisclosed and iceberg orders. Undisclosed orders must have a volume in excess of $500,000, and appear in the screen with the volume field showing '/u.' Given that these orders have a degree of pre-trade transparency we do not consider them in our analysis. Iceberg orders display only a fraction of the order volume. The hidden portion ranks behind all displayed orders at the same price. We are unable to observe iceberg orders in the data; however, anecdotally, these orders are believed to be infrequent.

[10] The Online Appendix is available at http://goo.gl/yIaOJR.

changed to remove the 'ten-second rule,' which had required brokers to place an order in the CLOB for ten seconds before executing a dark trade. This change made it easier for brokers to execute dark trades, especially when using algorithms. On June 28, 2010, the ASX Operating Rules were amended to allow dark trades to be executed at the midpoint of the best bid and ask price, as well as at the best prices.[11] On July 1, 2010, the ASX reduced its trading fees for all trade types, with a larger fee decrease for CLOB trades than for block and dark trades. Further details about these changes are provided in the Online Appendix.

There is one important difference between the US and Australian markets that is worth noting for readers unfamiliar with the Australian market. Unlike the US market, in which almost all marketable retail order flow is routed to wholesale market makers,[12] in Australia, retail order flow is almost exclusively executed on the ASX.[13] This is likely influenced by the fact that payment for order flow is not permitted in Australia. This difference means that the dark order flow examined in this paper is more similar to the dark order flow executed in US dark pools, rather than the dark order flow executed by US wholesale market makers.

4. Data and descriptive statistics

Our sample comprises the constituents of the All Ordinaries Index, which includes the 500 largest (by market capitalization) ASX-listed stocks and accounts for over 95% of the total market capitalization. Our sample period extends from February 1, 2008 to October 30, 2011. The end of the sample period is chosen to avoid confounding effects from fragmentation in lit liquidity resulting from the launch of a second lit exchange, Chi-X, on October 31, 2011.

We obtain millisecond-stamped data on all trades and all CLOB and Centre Point orders (including order entry, amendment, and cancellation messages) for our sample from the AusEquities database maintained by the Securities Industry Research Centre of Asia-Pacific. During our sample period, all trades are required to be reported to the

---

[11] Unlike the US, Australian dark pools can only trade at the midpoint or on the same price grid as the exchange. Therefore, the competition based on tick size discussed in Kwan et al. (2015) is not relevant in the Australian market.

[12] See SEC Concept Release on Equity Market Structure (2010) for further details.

[13] Australian Securities and Investments Commission (ASIC) (2013) estimates that in September 2010, only 4% of retail order flow is executed away from the exchange.

exchange immediately. As a result, we have a single consolidated source for all trade types: lit, dark, and block trades. Therefore, we minimize issues which arise in the US and other markets due to inconsistencies in time-stamps across different trading venues and inaccuracies with classification of dark and lit trade types.[14] For most of our analysis, we restrict our sample to the ASX continuous trading hours of approximately 10:00 to 16:12.[15]

We precisely classify CLOB trades and Centre Point trades as buyer- or seller-initiated by tracing trades back to their originating orders using the order identifiers recorded in the data. We define the trade initiator as the counterparty to a trade that submits their order last. In cases where we do not observe the original orders (block trades and dark trades other than those executed on Centre Point), we classify trades as buyer- (seller-) initiated if the trade price is above (below) the prevailing CLOB midquote.[16]

Fig. 1, Panel A, provides a time-series of dark, block, and total trading on the ASX over the months February 2008 to October 2011. The combined proportion of dollar volume executed using dark and block trades is approximately 18% and does not exhibit a clear trend over the period. However, from early 2010, there is an upward trend in the share of dark trades and a downward trend in the share of block trades. This reflects the growth in the number of dark pools. It also indicates that dark trades might be used as a substitute for block trades as brokers increase their use of algorithms.

< Insert Fig. 1 here >

Fig. 1, Panel B, shows that average trade sizes decline substantially over our sample period for all trade types, although the rate of decline is greatest for dark trades.

---

[14] For example, in the US the Trade Reporting Facility (TRF) is often used to proxy for dark trades. However, the TRF includes trade reports for ECNs, which are lit trading venues. Prior to BATS and Direct Edge being registered as exchanges, these lit venues accounted for substantial fractions of TRF volume. The misclassification also applies to lit trades, as exchange volumes include executions resulting from the execution of hidden orders on exchange. SEC statistics indicate that these account for around 10% to 12% of exchange volume, with substantial variation across exchanges.

[15] Opening call auctions take place at a random time within a 30-second window, and stocks commence trading in batches between 10:00 and 10:09. Closing call auctions take place in a single batch between 16:10 and 16:12 at a random time within a 60-second window.

[16] We are unable to reliably assign the trade initiator for midquote trades from dark systems other than Centre Point. Such trades constitute less than 1% of all dark trades (less than 0.1% of all trades). We omit these trades in the analysis that requires trade direction (calculation of the information content of trades).

The average size of dark trades declines from approximately $150,000 to $10,000, likely due to the increasing use of algorithms to manage executions in dark pools. Similarly, the average size of lit trades declines from approximately $13,000 to $5,000, again due to increasing use of execution algorithms and growth in high-frequency trading.

Table 1 reports descriptive statistics for the stocks in our sample. The average stock-day has 1,050 trades with a total value of $9.91 million. The median stock-day has substantially fewer trades with around 270 with a total value of $0.7 million. Table 1 also reports that the average (median) company in the sample has a market capitalization of $2.75 billion (422 million). The average spread of 129 basis points (bps) is considerably higher than the median spread of 67 bps. Price displays a similar pattern with an average of $4.76 and median of $1.79. On average, approximately 60% of the time stocks trade at the minimum possible spread of one tick ($0.01 for stock prices greater than $2). An average stock-day has around 4.6 quote messages per trade.

< Insert Table 1 here >

Table 2 reports correlations between the explanatory variables used in our analysis. Larger stocks and high volume stocks/days tend to have higher levels of dark and block trading. The levels of dark and block trading are negatively correlated with bid-ask spreads and volatility. These results are consistent with the cross-sectional and time-series variation in the level of activity in US dark pools reported by Buti et al. (2011), suggesting that the use of dark trading in Australia is similar to that in the US. Unsurprisingly, large stocks tend to have higher volume, lower spreads, and lower volatility. High volume is associated with low spreads and low volatility. Spreads are more often constrained by the tick size in large stocks, when volume is high, and the spread is relatively narrow. Stocks with constrained spreads also tend to have a higher level of dark trading, consistent with Kwan et al. (2015) who show that dark trading is often used as a way of obtaining a finer pricing grid than what is allowed on a lit market.

< Insert Table 2 here >

13

5. Empirical approach

Our empirical approach involves: (i) estimating a variety of price discovery characteristics for each stock-day in our sample using intraday data; and (ii) relating the price discovery characteristics to dark and block trading via stock-day panel regressions. For the panel regressions we take two approaches: (i) one-stage OLS regression with a range of control variables and fixed effects (stock and time); and (ii) two-stage least squares (2SLS) instrumental variables (IV) regressions.

While the IV models have the advantage of explicitly addressing the potential endogeneity of dark trading, we also report the one-stage ordinary least squares (OLS) regression results for several reasons.[17] First, anecdotal accounts of how traders choose to execute trades indicate that endogeneity concerns are likely to be more severe in causally relating dark trading and liquidity, than relating dark trading and price discovery. Our empirical results support this view; the results from our IV models are qualitatively similar to the one-stage OLS results (in some cases stronger in magnitude, suggesting that endogeneity might act against us finding a significant result). Second, the one-stage OLS regression models are simpler and likely to have higher statistical power. Therefore, if the *potential* for endogeneity does not ultimately have a large impact on our estimates, the one-stage OLS regression models could be preferable due to their higher precision.

The general form of our linear panel regressions is:

$$y_{id} = \alpha + \beta_{DARK} DARK_{id} + \beta_{BLOCK} BLOCK_{id} + \sum_{j=1}^{6} \delta_j C_{jid} + \varepsilon_{id}, \qquad (1)$$

where y_{id} is one of the price discovery characteristics for stock i on day d, and C_{jid} is a set of j control variables including log market capitalization, log bid-ask spread, the proportion of the trading day for which the stock's spread is constrained to one tick, log total dollar volume, volatility (standard deviation of one-minute midquote returns), and the messages-to-trades ratio, which serves as a proxy for algorithmic trading.[18] $DARK_{id}$

---

[17] The form of endogeneity that could be a concern is traders conditioning their order submission strategies (whether to send an order to the dark or to the lit market, which in turn determines the share of dark trading) on the price discovery characteristics. For example, Buti, Rindi, and Werner (2013) show that orders are more likely to be sent to a dark pool when liquidity is higher.

[18] Due to relatively high correlations between some control variables, in robustness tests we in turn exclude market capitalization, spread, and total dollar volume from the regressions. The results, including coefficients on other control variables, are very similar suggesting that multicolinearity does not pose a significant problem.

and $BLOCK_{id}$ measure the dollar volume of dark and block trades, respectively, as a percentage of the stock-day's total dollar volume.[19] In our 2SLS IV models $DARK_{id}$ and $BLOCK_{id}$ are replaced with fitted values from first-stage regressions. In the first-stage regressions, $DARK_{id}$ and $BLOCK_{id}$ are regressed on the set of instrumental variables and the control variables, for each stock. F-tests of the null hypothesis that the instruments do not enter the first-stage regression show that our tests do not suffer from the problem of weak instruments.[20]

We use two different sets of instruments for robustness, and find consistent results across the two sets. Our first set is based on market structure changes that are exogenous with respect to price discovery, but influence the amount of dark trading. This is similar to the approach used in other studies, such as Hendershott, Jones, and Menkveld (2011), except that our instruments contain only time-series variation because they affect all stocks at the same points in time. The market structure changes include the removal of the ten-second rule on November 30, 2009, which made it easier to execute dark trades. We construct a dummy variable ($D_t^{NO-10SEC-RULE}$) that takes the value one after the change, and zero before. The change in ASX trading fees on July 1, 2010 occurred largely in anticipation of competition from other market operators, and changed the relative explicit costs of trading in the dark compared to trading in the CLOB. Similarly, the launch of ASX's own dark pool, Centre Point, is unlikely to have been motivated by price discovery characteristics, but had an impact on the amount of dark trading. Because the change in trading fees and the launch of Centre Point took place three days apart, we construct a single dummy variable ($D_t^{NEW-FEES}$) that takes the value of one after both changes occurred, and zero before. Finally, the growth in the number of dark pools from four at the start of the sample period to 16 at the end has increased the ability to automate

---

[19] We use the share of dollar volume in our primary specification and in robustness tests we re-estimate the regressions using the share of trades instead of dollar volume. We also estimate the regressions using log-transformed $DARK_{id}$ and $BLOCK_{id}$ metrics, which reduce the influence of extreme values in dark and block volumes. The results are similar and our conclusions hold under both measures.

[20] Bound, Jaeger, and Baker (1995, p. 446) state that "F statistics close to 1 should be cause for concern." When instrumenting for $DARK_{id}$, the average F-statistics for the first and second set of instruments are 5.69 and 4.89, respectively, both well in excess of levels that warrant concern. When instrumenting for $BLOCK_{id}$, the average F-statistics for the first and second set of instruments are 2.96 and 65.58—the former being the lowest of the four F-statistics because the first set of instrumental variables is chosen specifically to instrument for the level of dark trading, not necessarily block trading.

dark executions. We construct an instrumental variable that measures the number of dark pools in operation, $DarkVenues_t$, as well as its square, $DarkVenues_t^2$.[21] Together these four variables form our first set of instruments. When using the first set of instruments we also control for a time trend in both the first- and second-stage regressions to minimize the possibility that the instruments pick up any general trends in dark and block trading.

For our second set of instruments we follow Hasbrouck and Saar (2013) and Buti et al. (2011) and instrument the level of dark trading in a stock-day with the average level of dark trading on that day in all other stocks in the corresponding size (market capitalization) quartile. This variable meets the requirements for an instrument because the level of dark trading in other stocks is correlated with the level of dark trading in a particular stock (95% confidence interval for the pooled Pearson correlation coefficient is 0.154 to 0.160), and dark trading in other stocks is unlikely to be driven by the nature of price discovery in the particular stock. Similarly, we instrument the level of block trading with the average level of block trading on that day in all other stocks in the corresponding size quartile. When using the second set of instruments we include an additional control variable in both the first- and second-stage regressions ($Y_{OtherStocks}$), which measures the average of the corresponding dependent variable on that day for all other stocks in the corresponding size quartile. This variable helps the instruments isolate the causal effect of dark and block trading, by removing potential reverse causality that can arise from cross-sectional commonality in market characteristics.

To investigate nonlinearity, we estimate an alternative version of Eq. (1) in which we replace the continuous variables $DARK_{id}$ and $BLOCK_{id}$ with a series of dummy variables that measure dark trading (D_{id}^{range}) and block trading (B_{id}^{range}) over various ranges:

$$
\begin{aligned}
y_{id} = \alpha &+ \beta_1 D_{id}^{0-5\%} + \beta_2 D_{id}^{5-10\%} + \beta_3 D_{id}^{10-20\%} + \beta_4 D_{id}^{20-30\%} + \beta_5 D_{id}^{30-40\%} + \beta_6 D_{id}^{>40\%} \\
&+ \gamma_1 B_{id}^{0-5\%} + \gamma_2 B_{id}^{5-10\%} + \gamma_3 B_{id}^{10-20\%} + \gamma_4 B_{id}^{20-30\%} + \gamma_5 B_{id}^{30-40\%} + \gamma_6 B_{id}^{>40\%} + \sum_{j=1}^{6} \delta_j C_{jid} + \varepsilon_{id}.
\end{aligned} \tag{2}
$$

---

[21] Our results are robust to different combinations of these variables, including omitting the variable $DarkVenues_t^2$.

The omitted, reference category corresponds to zero dark and zero block trading.[22] This specification is able to characterize many forms of nonlinearity that would be difficult to fit with a polynomial. We rely on this nonlinear specification to estimate the economic magnitudes of the associations between dark/block trading and price discovery.

6. Informativeness of trades and impact on spreads

We start by examining the informativeness of lit, dark, and block trades as this is one of the issues for which theory is conflicting. Whether dark trading segments informed and uninformed traders and which venue is used disproportionately by informed traders has important implications for liquidity and price discovery. Ye (2012) argues that when a dark venue co-exists with a lit venue, an informed trader will scale back the aggressiveness of his trading in the lit market to make larger profits in the dark. This means informed traders will execute a considerable share of their trades in the dark and therefore dark trades should be relatively informative. In contrast, Zhu (2014) predicts that informed traders face a lower execution probability in the dark compared to uninformed traders and therefore uninformed traders will execute a disproportionately higher share of their trades in the dark. It therefore follows from Zhu's model that dark trades should be less informative than lit trades. Dark trades could also be less informed than lit trades due to dark pools deliberately excluding certain types of relatively informed traders, or due to participants in the pool opting-out of executions with relatively informed traders.

There are also other reasons why both informed and uninformed traders might be attracted to relatively non-transparent trading venues. For uninformed traders the lack of transparency can help reduce "picking off" risks and exploitation by predatory traders, while for informed traders a lack of transparency can help prevent information leakage. Therefore, whether relatively more or less informed trades occur in the dark is an

---

[22] As an example of how the dummy variables are defined, $D_{id}^{0-5\%}$ takes the value 1 if the share of stock-day id's dollar volume executed in the dark is $0 < DARK_{id} \leq 5\%$, and zero otherwise. The dummy variable $B_{id}^{10-20\%}$ takes the value one if the share of stock-day id's dollar volume executed as block trades is $10\% < BLOCK_{id} \leq 20\%$. Therefore, the coefficients of the dummy variables estimate the effect of different levels of dark/block trading relative to the case of no dark/block trading. Our robustness tests indicate that the results are not sensitive to the choice of ranges.

empirical question, and one that has implications for how price discovery occurs and how adverse selection risk changes in response to dark trading.

To measure the informativeness of different trade types (lit compared to dark and block) we adapt the Hasbrouck (1991) vector auto-regression (VAR) framework to our trade-type partition. We calculate signed dollar volume of lit, dark, and block trades, x_t^{LIT}, x_t^{DARK}, and x_t^{BLOCK}, in every one-second interval, t, for every stock-day. For each stock-day we estimate the following system:

$$x_t^{LIT} = \mu^{LIT} + \sum_{i=1}^{60} \phi_i^r r_{t-i} + \sum_{i=1}^{60} \phi_i^{LIT} x_{t-i}^{LIT} + \sum_{i=1}^{60} \phi_i^{DARK} x_{t-i}^{DARK} + \sum_{i=1}^{60} \phi_i^{BLOCK} x_{t-i}^{BLOCK} + \varepsilon_t^{LIT}$$

$$x_t^{DARK} = \mu^{DARK} + \sum_{i=1}^{60} \theta_i^r r_{t-i} + \sum_{i=1}^{60} \theta_i^{LIT} x_{t-i}^{LIT} + \sum_{i=1}^{60} \theta_i^{DARK} x_{t-i}^{DARK} + \sum_{i=1}^{60} \theta_i^{BLOCK} x_{t-i}^{BLOCK} + \varepsilon_t^{DARK}$$

$$x_t^{BLOCK} = \mu^{BLOCK} + \sum_{i=1}^{60} \lambda_i^r r_{t-i} + \sum_{i=1}^{60} \lambda_i^{LIT} x_{t-i}^{LIT} + \sum_{i=1}^{60} \lambda_i^{DARK} x_{t-i}^{DARK} + \sum_{i=1}^{60} \lambda_i^{BLOCK} x_{t-i}^{BLOCK} + \varepsilon_t^{BLOCK} \quad (3)$$

$$r_t = \mu^r + \sum_{i=1}^{60} \gamma_i^r r_{t-i} + \sum_{i=0}^{60} \gamma_i^{LIT} x_{t-i}^{LIT} + \sum_{i=0}^{60} \gamma_i^{DARK} x_{t-i}^{DARK} + \sum_{i=0}^{60} \gamma_i^{BLOCK} x_{t-i}^{BLOCK} + \varepsilon_t^r,$$

where t indexes one-second intervals (individual stock and date subscripts are suppressed) and r_t is the log-midquote change in the tth interval.

After estimating the above system for each stock-day, we calculate the informativeness of lit, dark, and block volume as the cumulative impulse response (measured 60 seconds forward in time) of midquote returns for a shock of +$10,000 of signed lit, dark, and block volume, respectively, holding all other variables equal to their unconditional means. Following Hasbrouck (1991) we interpret the permanent price impact of order flow as a measure of the private information contained in the order flow. To minimize the effects of outliers we winsorize the permanent price impact measures by setting extreme positive and negative values to the 1st and 99th percentile values, for each stock and each date.

Table 3 reports permanent price impacts of lit, dark, and block trades. The average permanent price impact of lit trades is larger than that of dark trades (3.62 bps per $10,000 for lit trades versus 3.31 bps for dark trades with equal-weighting of stock-days, and 0.65 bps for lit versus 0.41 bps for dark with dollar volume-weighting). The difference in lit and dark price impacts is statistically distinguishable at the 1% level using paired t-tests. The average permanent price impacts of lit and dark trades are both higher than the average permanent price impact of block trades (0.15 bps with equal-

18

weighting and 0.02 bps with dollar volume-weighting) and these differences are also statistically significant at the 1% level.[23]   The median permanent price impact for lit trades (1.91 bps) is considerably higher than the medians for dark and block trades (0.03 bps and 0.01 bps) and the difference in medians is statistically significant at the 1% level using paired sign tests.   Therefore, a 'typical' (median) lit trade contains considerably more private information per unit of volume than dark and block trades.   On average, dark and block trades do contain some information, and in particular, some dark trades are highly informed.

< Insert Table 3 here >

These results are consistent with the predictions of Zhu (2014) that a relatively larger proportion of uninformed trades will execute in the dark because they are less likely to cluster on one side of the market compared to informed trades.[24]   Zhu predicts that the higher proportion of uninformed trades in the dark will leave behind a higher concentration of informed traders in the lit market, which will result in higher adverse selection risk and wider spreads in the lit market.   In contrast, Ye (2012) predicts that the informed trader will scale back their aggressiveness in the lit market, and consequently, one might expect lower adverse selection risk and narrower spreads in the lit market as a result of dark trading.   Therefore, to provide further evidence on the conflicting theoretical predictions and the mechanisms by which dark trading impacts markets, we examine how spreads in the lit market are affected by dark trading.

Consistent with the predictions of Zhu (2014) and the notion that a larger share of uninformed trades will execute in the dark, Table 4 indicates that quoted spreads become wider in the lit limit order book as dark and block trading increase.   The impact of dark

---

[23] The price impacts presented in this section are all per $10,000 of volume.   Because block trades are much larger than lit trades, the total price impact of a block trade is larger than the total price impact of a lit trade.   The relatively low informativeness of block trades (per unit volume) is consistent with previous studies that find upstairs markets tend to be used by traders who can credibly signal that their trades are uninformed (e.g., Madhavan and Cheng, 1997; Bessembinder and Venkataraman, 2004; Booth, Lin, Martikainen, and Tse, 2002).

[24] An alternative explanation why dark trades tend to be less informed compared to lit trades is that dark trading venues are more selective in what order flow they accept and thereby screen out some informed traders.   For example, dark pools have selective membership whereby only relatively uninformed traders are allowed access (see Boni et al., 2012), and brokers that know their clients are able to be selective in what order flow they internalize and what they send to a lit market for execution.   These explanations are not mutually exclusive.

trading on spreads in the lit market is highly statistically significant across all of our regression specifications: one-stage OLS without fixed effects, with stock fixed effects, with date fixed effects, and 2SLS IV regressions using two different sets of instruments. We find similar results using spread measured in cents, rather than spread expressed in basis points. The control variables indicate that spreads tend to be narrower for larger stocks and during higher trading activity. High volatility tends to be associated with wider spreads. Average spreads have declined through time and have commonality across stocks (positive coefficient on $Y_{OtherStocks}$).

< Insert Table 4 here >

The magnitude of the increase in quoted spreads is also economically meaningful. For example, estimates using the one-stage OLS regression model with dummy variables for different levels of dark and block trading (Eq. (2)) show that increasing dark trading from zero to 10% of dollar volume is expected to increase quoted spreads by 11% after controlling for other factors. This means that for the average stock, spreads will increase from 128 bps to 142 bps. A more modest increase in dark trading from 10% to 12.5% is expected to increase spreads by 2.2% (an increase of 2.8 bps for the average stock). The 2SLS IV estimates are even larger in magnitude. For example, an increase in dark trading from 10% to 12.5% is expected to increase spreads by 6.5% to 7.2%, depending on which set of instruments is used (an increase of 8.4 bps to 9.3 bps for the average stock). Similarly, an increase in block trading from 10% to 12.5% of dollar volume is expected to increase spreads by 1.7% using the one-stage OLS model and 1.7% to 4.0% using the 2SLS IV models with two different sets of instruments. Wider spreads increase the costs of trading in the lit market, which can encourage order flow to migrate away from the lit market in a self-reinforcing spiral.

To understand why our results on the informativeness of trades and impacts on adverse selection are consistent with Zhu (2014) and differ from Ye (2012), it is useful to consider the underlying assumptions in these models. An important assumption in Ye (2012) is that the informed trader is a monopolist (as in the standard Kyle model), whereas in Zhu (2014) there are many competing informed traders. As a consequence of being a monopolist, the informed trader in Ye's model does not face the problem of low

execution probability in the dark, which occurs in Zhu's model due to the tendency for informed traders to cluster on one side of the market. Low execution probability dissuades informed traders from the dark in Zhu's model making dark trades relatively uninformative. A further consequence of the monopolist assumption in Ye's model is that the informed trader is able to scale back the aggressiveness of his orders in the lit market. In contrast, competition among informed traders in Zhu (2014) severely limits opportunities for scaling back aggressiveness. The empirical support for Zhu's model suggests that perhaps the assumption of multiple competing informed traders is more realistic in our sample than the assumption of a monopolistic informed trader.

7. Price discovery shares

Results in the previous section suggest that, consistent with Zhu (2014), less informed order flow tends to execute in the dark and that dark trading increases adverse selection risks in the lit market. To provide further insights about these processes, in this section we analyze how and where information enters the market, and to what extent dark and block trade prices contribute to price discovery.

Theory provides two predictions. First, if a random selection of trades were to migrate from the lit market to the dark, the dark market's share of price discovery should increase roughly in line with the increase in the dark market's share of trading activity; a 1% increase in the dark market share should lead to an increase of approximately 1% in the dark's share of price discovery. If, however, the trades that migrate to the dark tend to be less informed than average trades, as predicted by Zhu (2014), then the dark market's share of price discovery should increase at a rate less than the increase in the dark market's share of trading activity.

Second, informed traders face a comparative advantage in providing liquidity when adverse selection risks are high because of their informational advantage (e.g., Rindi, 2008). Therefore, if dark trading increases adverse selection risks in the lit market, we would expect an increase in the propensity for informed traders to supply liquidity, and thus an increase in the amount of price discovery that takes place through quotes relative to that of trades.[25]

---

[25] Goettler, Parlour, and Rajan (2009) point out that the informativeness of the best quotes relative to the informativeness of trade prices depends on the extent to which informed traders supply liquidity by trading

21

Two traditional approaches to measuring the contributions of different markets or types of order flow to price discovery are Hasbrouck's (1995) information share (*IS*) and Gonzalo and Granger's (1995) common factor share (*CS*). Fundamentally, both methods decompose price innovations into permanent and temporary components. As pointed out by Yan and Zivot (2010) and Putniņš (2013), both metrics measure (with different weights) a combination of two dimensions of market efficiency: (i) timeliness in impounding of new information; and (ii) avoidance of transitory shocks. For the purpose of identifying where information enters the market, we are interested in measuring the first component: the extent to which a price or type of order flow is the first to impound new information about the 'true' underlying asset value. Of the two traditional measures Hasbrouck's *IS* comes closer to identifying the leader in impounding new information, but is also influenced by the relative amount of noise in the price channels (Putniņš, 2013). To isolate the relative speed at which information is impounded by a price series from its relative level of noise, Putniņš (2013) extends the analytic results of Yan and Zivot (2010) and defines the "information leadership share" (*ILS*) as:

$$ILS_1 = \frac{\left| \dfrac{IS_1}{IS_2} \dfrac{CS_2}{CS_1} \right|}{\left| \dfrac{IS_1}{IS_2} \dfrac{CS_2}{CS_1} \right| + \left| \dfrac{IS_2}{IS_1} \dfrac{CS_1}{CS_2} \right|}, \quad ILS_2 = \frac{\left| \dfrac{IS_2}{IS_1} \dfrac{CS_1}{CS_2} \right|}{\left| \dfrac{IS_1}{IS_2} \dfrac{CS_2}{CS_1} \right| + \left| \dfrac{IS_2}{IS_1} \dfrac{CS_1}{CS_2} \right|}. \tag{4}$$

The ILS_1 and ILS_2 each have the range [0,1] (and they sum to one), similar to *IS* and *CS*, with values above (below) 0.5 indicating the price series impounds new information faster (slower) than the other price series and thereby leads (does not lead) the process of price discovery. Using simulations, Putniņš (2013) shows that *ILS* is robust to differences in noise levels and therefore correctly attributes price discovery in a wider range of settings. Therefore, we report results using *ILS* (results using *IS* are available upon request).

Following Hasbrouck (1995), we estimate the following vector error correction model (VECM) for each stock-day using one-second intervals, *t*:

---

with limit orders versus demand liquidity with market orders. If informed traders tend to demand liquidity, then trade prices will convey relatively more information than quotes. In contrast, if informed traders tend to supply liquidity, quotes will be relatively informative compared to trade prices.

$$\Delta p_{1,t} = \alpha_1 (p_{1,t-1} - p_{2,t-1}) + \sum_{i=1}^{60} \gamma_i \Delta p_{1,t-i} + \sum_{j=1}^{60} \delta_j \Delta p_{2,t-j} + \varepsilon_{1,t}$$

$$\Delta p_{2,t} = \alpha_2 (p_{1,t-1} - p_{2,t-1}) + \sum_{k=1}^{60} \phi_k \Delta p_{1,t-k} + \sum_{m=1}^{60} \varphi_m \Delta p_{2,t-m} + \varepsilon_{2,t},$$

(5)

where $p_{1,t}$ and $p_{2,t}$ are the last available log prices of price series 1 and 2, respectively.[26] We calculate IS_1, IS_2 and CS_1, CS_2 from the error correction parameters and variance-covariance of the error terms, following Baillie, Booth, Tse, and Zabotina (2002), and ILS_1 and ILS_2 following Putniņš (2013).

We estimate two different versions of the VECM above. In the first, the two price series are: (i) lit trade prices; and (ii) dark/block trade prices. This version allows us to analyze the relative contribution to price discovery made by lit trades compared to dark and block trades. In the second, the two price series are: (i) midquotes, calculated from the prevailing best bid and ask prices; and (ii) trade prices, using the last available trade price irrespective of the trade type. This version allows us to analyze the contribution of the best quotes, compared to trade prices.

The price discovery share of lit trades, compared to dark/block trades, indicates that lit trades contribute substantially more to price discovery (median ILS_{LIT} of 0.84).[27] This is because lit trades account for a larger share of volume than dark and block trades but also because typically, lit trades are more informative. Table 5 reports panel regression results of how the price discovery share of lit trades changes as dark and block trading increase. As the share of block trading increases, lit trades contribute relatively less to price discovery, compared to dark and block trades. The impact of dark trading is less clear, because the coefficient on the variable *DARK* is only statistically distinguishable from zero in two of the five regression specifications (in those specifications the coefficient is negative, similar to the effect of block trades).

---

[26] Estimation of the information share metrics relies on price series being co-integrated. In studies of cross-listed stocks the law of one price keeps the two prices of the stock within certain arbitrage limits and therefore ensures co-integration. In this paper, we study the contribution of limit order book quotes as well as prices of different types of trades (lit and dark/block) for each stock within one market, similar to Anand and Subrahmanyam (2008). The limit order book quotes, lit trade prices, and dark/block trade prices for a stock are all linked to the fundamental value of the stock and are therefore co-integrated.

[27] The median of ILS_{LIT} does not take into consideration the stock-days on which we are unable to compute the information shares of lit and dark/block trade prices (days when there are zero or very few dark/block trades). On such days, dark/block trades make little or no contribution to price discovery. Therefore, ILS_{LIT} understates the contribution of lit trades to price discovery overall.

23

< Insert Table 5 here >

As the share of volume executed in the dark (or executed as blocks) increases, the price discovery share of lit trades should decrease somewhat mechanistically (Anand and Subrahmanyam, 2008). If the price discovery shares changed proportionally to the volume shares, we would expect the coefficients on *DARK* and *BLOCK* to be around -1, indicating that an *x%* increase in the share of dark or block volume leads to an *x%* decrease (increase) in the share of price discovery attributable to lit trades (dark/block trades). However, in the regressions of *ILS$_{LIT}$* (Table 5), the coefficients on *DARK* and *BLOCK* across the five specifications range from -0.287 to -0.004. Therefore, as dark and block trading increase, their contribution to price discovery increases at a slower rate than their volume share. This indicates dark and block trades on average contain less private information than lit trades consistent with our previous results and consistent with the predictions of Zhu (2014).

Turning to the contribution of lit quotes to price discovery, the median information leadership share of midquotes (*ILS$_{MIDQUOTE}$*) is 0.56, indicating that overall, lit quotes make a slightly larger contribution to impounding new information than trades. This is consistent with other studies that show informed traders often trade with limit orders and provide liquidity.[28] Table 5 shows that an increasing share of dark trading is associated with an increase in the *ILS* of the midquote, holding other variables fixed. This is true in the single-stage OLS specifications as well as the 2SLS IV specifications. The effect of block trading is opposite, consistent with the fact that block trade prices are not mechanically derived from the prices in the lit market unlike the prices of many dark trades, and block trades tend to be less informed.

The tendency for midquotes to become relatively more informative (suggesting that informed traders increasingly supply liquidity in the lit market) as the share of dark trading increases is consistent with our earlier results on how dark trading impacts adverse selection. As Rindi (2008) and others point out, informed traders are particularly effective liquidity suppliers when adverse selection risks are high because of their informational advantage. Therefore, the increasing informativeness of midquotes is

---

[28] For example, Bloomfield, O'Hara, and Saar (2005), Kaniel and Liu (2006), Goettler et al. (2009), Roşu (2013), and Boulatov and George (2013).

consistent with increased adverse selection risk in the lit market as a disproportionately large share of uninformed trades execute in the dark.

8. Informational efficiency

The previous sections show that dark venues segregate traders: less informed traders migrate to the dark, while more informed traders tend to concentrate in the lit market. The segregation of traders not only impacts the liquidity of the lit market, but also affects informational efficiency by changing the incentives to become informed.

Theory suggests that segregation of informed and uninformed traders could harm or benefit informational efficiency depending on the nature of private information. When uninformed traders migrate to the dark, the profitability of acquiring information decreases because informed traders are less able to trade in the dark (because they tend to cluster on one side of the market and/or because of restrictions by dark pool operators). Thus, fewer traders choose to become informed and/or informed traders acquire less costly, less precise information. If traders that choose to become informed receive unique noisy signals (e.g., the fundamental value plus an independent error), decreased profitability of acquiring information leads to less information production in aggregate and therefore less informative prices (e.g., Kyle, 1981, 1984, 1989; Admati and Pfleiderer, 1988). In such circumstances, the segregation caused by dark trading harms price discovery.[29]

In contrast, if all traders that choose to become informed receive an identical piece of information as in Zhu's (2014) model (e.g., exact knowledge of the fundamental value), the segregation caused by dark trading can improve price discovery. This is because a decrease in the number of informed traders (due to fewer uninformed traders in the lit market) corresponds to a decrease in the degree of competition on the same set of private information, but no change in the amount of private information that in aggregate is held by informed traders. In such a case, a higher concentration of informed traders in the lit market improves the market's ability to incorporate information from order flow and thus improves price discovery. Therefore, whether dark trading increases or decreases the informativeness of prices is ultimately an empirical question.

---

[29] Another reason that dark trading could harm price discovery is that the lower level of transparency (e.g., inability to observe dark orders that do not execute) impedes the market's ability to incorporate information from order flow into prices.

To examine this question, we use three types of informational efficiency measures commonly used in empirical studies: (i) absolute values of midquote return autocorrelations at various horizons; (ii) absolute values of excess variance ratios; and (iii) measures of short-term return predictability using lagged market returns. We calculate the informational efficiency measures each stock-day using intraday data. Details of the calculation and interpretation of these measures are provided in Appendix A. It is worth noting that while the return predictability will effectively capture permanent price movements toward fundamental value, the autocorrelation and variance ratio measures will pick up temporary as well as permanent price movements. When measured at a high frequency, autocorrelations and variance ratios are likely to be significantly influenced by temporary price movements and thus pick up illiquidity as well as informational efficiency. Therefore, we also consider low-frequency informational efficiency measures, which will be less prone to this issue.

Using high-frequency informational efficiency metrics is important to maximize the statistical power of our tests. Rösch, Subrahmanyam, and van Dijk (2013) provide evidence that such informational efficiency metrics measured at intraday horizons are highly correlated with low-frequency measures of informational efficiency and are different from liquidity measures. Anderson, Eom, Hahn, and Park (2013) find that partial price adjustment (slow price adjustment and overshooting), which implies a degree of informational inefficiency, is a major source of positive and negative autocorrelations. In robustness tests we confirm that our results hold at lower frequencies (estimating the measures for each stock-month using daily data), although such tests have lower statistical power and less precision.

Table 6 reports panel regression estimates of the relation between the share of dark/block trading and informational efficiency, using one-stage OLS. The general pattern that emerges is that an increase in the share of dark trading, all else equal, is associated with deterioration in informational efficiency. The coefficients of $DARK_{id}$ are positive for all measures of informational inefficiency, and statistically significant for all specifications with and without fixed effects. The R^2s, which exclude the variation explained by the fixed effects, tend to be lower for specifications that include stock fixed effects than time fixed effects, indicating that the explanatory variables are able to explain a greater fraction of the cross-sectional variation in informational efficiency than

the time-series variation. The control variables indicate that larger stocks tend to be more informationally efficient. The relation between informational efficiency and spreads, volume, and volatility (after controlling for market capitalization) is not consistent across the different informational efficiency measures. The message-to-trade ratio is not statistically significantly related to efficiency. Overall, the results in Table 6 provide initial evidence that informational efficiency is harmed by dark trading.

< Insert Table 6 here >

Block trading, however, according to our evidence is not detrimental to informational efficiency. For all three informational efficiency measures the coefficients of $BLOCK_{id}$ are negative and they are statistically significant for two of the three measures, with and without fixed effects. This provides evidence that trading large blocks off-exchange can even be beneficial to the efficiency of the lit market. Previous studies identify two possible reasons why block trades can have a different effect on the market compared to dark trades. First, through the unique role of upstairs brokers as 'information repositories,' block trades are able to tap into additional liquidity that would not otherwise be expressed in the limit order book, thereby expanding aggregate liquidity (Grossman, 1992; Bessembinder and Venkataraman, 2004). Second, block trades are largely uninformed, but due to their size they would cause significant temporary price distortions if submitted to the limit order book without being broken-up into smaller trades. By being able to credibly signal the likely motivation for the trade in the upstairs market, a block trade's counterparty faces lower adverse selection risk, allowing the block trade to occur with a smaller price impact (Bessembinder and Venkataraman, 2004). These results are consistent with information provided by institutional brokers and investors who suggest the negotiation reveals valuable information about the nature and size of available liquidity, and sometimes identifies latent liquidity.

Turning to the 2SLS IV models of the impact of dark and block trading on informational efficiency (reported in Table 7), we find similar results. Across all three informational inefficiency measures, dark trading is associated with a statistically significant deterioration in informational efficiency, whereas block trading is estimated to have the opposite effect. These results have a similar level of statistical significance as

the one-stage OLS models and tend to have larger impacts. This provides evidence that our results relating dark trading to deterioration in informational efficiency are not driven by traders choosing to execute in the dark when informational efficiency is poor. If anything, the magnitudes suggest that endogeneity could work against finding significant results in our OLS regressions. These results provide evidence of a causal link from dark trading to deterioration of informational efficiency. The positive and highly significant coefficients on the control variable $Y_{OtherStocks}$ show significant cross-sectional commonality in informational efficiency. For two of the three efficiency measures the time trend is insignificant and for the third it is negative and marginally significant. We obtain similar results in specifications that omit the control variables.

< Insert Table 7 here >

The autocorrelations and variance ratios at the different frequencies (the components of $Autocorrelation_{Factor}$ and $VarianceRatio_{Factor}$) provide results that are consistent with those using $Autocorrelation_{Factor}$ and $VarianceRatio_{Factor}$. We also examine the autocorrelations and variance ratios without the absolute value transformation, i.e., allowing them to take positive and negative values. In our sample the autocorrelations and variance ratios tend to be negative—the pooled means are statistically significantly negative (ranging from -0.04 to -0.15) and even the 75th percentile values are negative for all of the different frequencies. The estimated effect of dark trading is to make the autocorrelations and variance ratios more negative (these results are not tabulated), consistent with a decrease in informational efficiency. An interpretation of these results is that prices (midquotes) tend to overreact to new information or order flow and subsequently reverse the overreaction (Anderson et al., 2013), and high levels of dark trading tend to exacerbate the inefficient overreactions and reversals. Negative autocorrelations also arise from imperfect risk-bearing capacity of liquidity providers (e.g., Ho and Stoll, 1981). Thus, some of the decrease in informational efficiency associated with high levels of dark trading could be due to decreases in liquidity.

The advantage of using relatively high-frequency measures of informational efficiency is high statistical power and the ability to provide a more granular

28

characterization of relations, such as the analysis of nonlinearity which we turn to next. The disadvantage is that the high-frequency measures are likely to capture illiquidity as well as informational inefficiency. Including the spread as a control variable helps isolate the impact on informational efficiency, distinct from the effect on liquidity. To further investigate this issue, we re-estimate the regressions using low-frequency variations of the informational inefficiency measures.[30]

Table 8 reports the results of 2SLS IV regressions using low-frequency informational efficiency measures. Similar to the high-frequency measures, dark trading tends to increase all three measures (the increases are statistically significant), suggesting deterioration of informational efficiency. Compared to the high-frequency efficiency measures, the magnitude and statistical significance of the impact of dark trading is lower for some measures (autocorrelations and variance ratios) but higher for others (*Delay*). Consistent with the low-frequency measures, block trading is not clearly detrimental to informational efficiency; in most regressions block trading has an insignificant effect, and the statistically significant effects are mixed. While these additional tests do not eliminate the fact that the informational inefficiency metrics are likely to also measure illiquidity to some extent, they do provide further evidence that dark trading impacts informational efficiency distinct from its impact on liquidity.

< Insert Table 8 here >

The effects of dark and block trading on informational efficiency could be nonlinear in their share of volume. For example, Eom et al. (2007) argue that market quality is an increasing concave function of transparency. This implies that an increase in dark trading from a low level is likely to have a smaller effect on market quality (and could even improve market quality) than the same magnitude increase from a relatively high level of dark trading. To investigate this possibility, we estimate nonlinear regression models given in Eq. (2) in which we replace the continuous variables $DARK_{id}$ and $BLOCK_{id}$ with a series of dummy variables that measure dark trading (D_{id}^{range}) and

---

[30] The low-frequency measures are computed every stock-month. We measure autocorrelations of one-day, two-day, and three-day returns, and then, like the high-frequency measures, we take the absolute value and first principal component. We measure absolute variance ratios that compare the variance of one-day returns to two-day, three-day and four-day return variance. We measure *Delay* in impounding market-wide information using regressions of each stock's daily returns on ten lags of market returns.

block trading (B_{id}^{range}) over various ranges. The patterns are similar when we use one-stage OLS and 2SLS IV approaches, and, therefore, to save space we only report results using the one-stage OLS approach.

Fig. 2 plots the coefficients of the dummy variables for each of the three informational efficiency metrics together with error bounds corresponding to +/- two standard errors.[31] Panel A suggests that low levels of dark trading are not harmful to informational efficiency (they could even be beneficial), but as dark trading increases it eventually reaches a 'tipping point' after which it has a negative impact. Specifically, after controlling for other stock characteristics, when dark trading accounts for approximately 10% of total dollar volume its impact on informational efficiency is very close to zero, i.e., it neither harms nor benefits informational efficiency. However, levels of dark trading above 10% of dollar volume are associated with lower informational efficiency compared to zero dark trading.

To illustrate the economic significance, an increase in dark trading from 10% to 20% of dollar volume is estimated to increase the informational inefficiency measures by 10% to 15% of a standard deviation using the one-stage OLS model, and 19% to 26% using the 2SLS IV models. A more modest increase in dark trading from 10% to 12.5% of dollar volume is expected to increase the informational inefficiency measures by 2% to 4% of a standard deviation using the one-stage OLS model, and 6% to 7% using the 2SLS IV models.

< Insert Fig. 2 here >

Fig. 2, Panel B, provides evidence that executing block trades away from the CLOB improves informational efficiency, but only up to a certain point. Maximum informational efficiency occurs around the point at which block trades account for approximately 15% of total dollar volume. Beyond this level, additional block trades tend to have a negative marginal impact on informational efficiency, although the total impact on informational efficiency remains positive until block trades account for approximately 40% of total dollar volume. Block trading at 15% of dollar volume is

---

[31] The range covered by each dummy variable is reduced to a single point for the purpose of the plots by taking the mean of $DARK_{id}$ and $BLOCK_{id}$ for the stock-days that fall into the corresponding range. For example, for stock-days that have dark dollar volume greater than zero but less than or equal to 5%, the mean of $DARK_{id}$ is 1.7%. Therefore, $D_{id}^{0-5\%}$ is plotted at the horizontal axis value of $DARK_{id} = 1.7\%$.

associated with improvements in the informational efficiency measures of approximately 14% to 21% of a standard deviation using the one-stage OLS models, and 5% to 14% using the 2SLS IV models. In general, small amounts of block trading away from the lit market are good for informational efficiency, but as with dark trading: too much can be harmful.

We also analyze nonlinearity in the impact of dark trading on the other variables: spreads and price discovery shares. The results (reported and discussed in the Online Appendix) show similar patterns to those for the informational efficiency measures, namely, that the marginal effects of dark/block trading at low levels are different to the effects at high levels. Low levels of dark trading are benign or even beneficial to spreads and adverse selection risks in the lit market, but high levels are harmful. The 'tipping point' beyond which market quality is harmed is lower for dark trading than for block trading.

A caution is in order. Our results do not indicate that dark trading in general is harmful; rather, they indicate that high levels of dark trading tend to harm price discovery, while low levels are benign or could even benefit price discovery. This distinction is important from a policy perspective—eliminating dark trading is unlikely to be an optimal regulatory response. Instead, our results suggest that regulators should seek to limit growth in dark trading beyond the point at which it becomes harmful. The distinction is also important in relating our analysis to the predictions of Zhu's (2014) model. The model predicts that in equilibrium adding a dark crossing system alongside a lit exchange will improve price discovery. High levels of dark trading are not necessarily equilibrium levels. Therefore, it is interesting to examine for how many stocks the current levels of dark trading are harmful to price discovery.

During the last ten months of our sample (January-October 2011), the median level of dark trading as a share of dollar volume was greater than 10% for 62 of the 498 stocks (12% of stocks). This illustrates that approximately 12% of stocks had levels of dark trading that were harmful to price discovery on most (>50%) trading days during the first ten months of 2011. On average, these stocks are larger, more actively traded, and more likely to have a constrained spread than the other stocks in the sample. Approximately one-third of the stocks in our sample have harmful levels of dark trading (>10% of dollar volume) on more than one-quarter of the trading days in 2011. No

stocks have block trading levels in excess of the 40% 'tipping point' for more than one-quarter of the trading days in 2011. Overall, these results demonstrate that block trading on a typical day is below harmful levels in all stocks, and dark trading on a typical day is below harmful levels for most stocks in our sample.

If the levels of dark trading in our sample reflect equilibrium levels, the calculations above suggest that Zhu's (2014) prediction that dark trading improves price discovery is violated for a significant fraction of stocks. One possible source of the disagreement between the model and our empirical results is modeling assumptions that are violated in practice. For example, as previously discussed, assumptions about the nature of private information are an important determinant of whether a decrease in the number of uninformed traders with which informed traders are able to interact harms or benefits price discovery. If informed traders endogenously acquire unique signals (rather than the same signal as in Zhu's model), then the segregation caused by dark trading could harm price discovery by causing less information production in aggregate.[32] Another possibility is that our empirical measures capture a different aspect of informational inefficiency than the inefficiency measured in the theoretical model. Similar to the way that liquidity has several dimensions that can be affected in opposite ways (e.g., spreads and depth), informational efficiency also has multiple dimensions that are not necessarily impacted in the same way.[33]

The calculations above also illustrate an important point: the 'tipping points' suggested by our analysis correspond to stock-day levels, and, therefore, should not be compared to market-wide aggregates of dark trading. The tendency for high-volume stock-days to have higher levels of dark trading means that market-wide aggregates of dark trading (effectively, volume-weighted averages) are typically higher than the median and the equal-weighted mean levels of dark trading.

---

[32] Other modeling assumptions that differ from practice include the trade execution rules (dark trades execute only at the midquote in the model, but at several prices in practice), market structure (a single dark venue in the model, but multiple dark venues in practice), the order types (in the model, informed and uninformed traders can only use market orders and not limit orders), lack of dynamics in the model, the fundamental value process, and so on.

[33] For example, one could measure the speed with which new information is reflected in prices or the precision (relative absence of noise) with which prices reflect the information. It is possible for speed to increase but at the expense of precision, or vice versa. Similarly, one can measure the extent to which prices reflect short-horizon information (past prices, prices of other stocks, order low, announcements), or long-horizon information (long-run cash flow forecasts, growth rates, discount rates). Again, it is possible for efficiency with respect to short-horizon information to increase and at the same time efficiency with respect to long-horizon information to decrease, or vice versa.

The tendency for less informed traders to trade in the dark (and in doing so avoid interacting with some informed traders) has important welfare implications. Although our results provide evidence that high levels of dark trading widen spreads in the lit market, this does not necessarily increase trading costs in aggregate because higher trading costs in the lit market could be offset by lower trading costs in the dark. As dark trading activity increases, costs associated with non-execution and delayed execution decrease in the dark. Although the impact on aggregate trading costs is not clear, dark trading leads to redistributions (transfers) of trading costs across different types of traders. The increase in trading costs in the lit market is largely borne by the informed traders that are less able to trade in the dark.

In the seminal models of Kyle (1985) and Glosten and Milgrom (1985), uninformed market makers break even on average and informed traders profit from trading on their information. As a result, uninformed traders on average lose an amount equal to the informed traders' profits. The wealth transfer from uninformed traders to informed traders occurs through the trading costs faced by uninformed traders: the bid-ask spread in Glosten and Milgrom (1985) and trade prices away from fundamental value in Kyle (1985). Importantly, the wealth transfer from uninformed traders to informed traders compensates them for the costs of producing information and thereby providing price discovery. In fact, when information acquisition is costly, the absence of uninformed ('noise') traders can cause a complete breakdown of price discovery resulting in an informationally inefficient market (Grossman and Stiglitz, 1980; Black, 1986). Therefore, the trading costs paid by uninformed traders play an important role in facilitating price discovery by compensating others for producing information. Our results are consistent with the notion that uninformed traders benefit from trading with each other in the dark, but their gain comes at the cost of less information production and therefore less informative prices.

9. Further robustness and subsample tests

In this section we detail a range of additional robustness tests. First, we examine whether our results hold for stocks of different sizes. We estimate our full set of analyses separately for large and small stocks (defined as market capitalization above and below the median, respectively). We find that our key results hold for both subsamples, in

particular, as dark trading increases informational efficiency deteriorates and spreads on the lit market increase. Large stocks also have a substantially lower proportion of stock-day observations with zero dark trading than small stocks. The consistency of results across the two groups therefore also provides some evidence of the robustness of the results to the proportion of zero dark trading observations included in the sample.

The results are similar in both the first and second halves of the sample period (2008–2009 and 2010–2011). This indicates that the potentially harmful effects of dark trading are not a new phenomenon. Given the changes in how dark trading takes place (increasing automation during the sample period due to an increasing number of dark pools), this result provides indirect evidence that the amount of dark trading matters for price discovery rather than the way in which dark trading takes place. Dark pools as such are not necessarily any more harmful than manual dark trading. However, if dark pools make it easier to trade in the dark, they could encourage growth of dark trading to levels that are harmful to price discovery.

Removing stock-days with prices above the 95th percentile and below the 5th percentile produces results that are very similar to those using the full sample. Similarly, the results are robust to excluding the control variables that are most correlated with other explanatory variables.

We examine alternative measures of dark/block trading activity, using number of trades instead of dollar volume, as well as log-transforms of the dark and block trading shares, and find similar results. Changes to the number of lags used in the VAR, VECM, and return predictability regressions do not have a substantial impact on our results. We also estimate the VAR and VECM models at lower frequency using ten-second intervals in place of one-second intervals, allowing the lags to span a ten-times larger window of past observations, and find qualitatively similar results.

Estimation of the VAR requires at least one lit trade, dark trade, and block trade, as well as changes in the midquote. This requirement is met and the VAR is successfully estimated for approximately 81,000 stock-days (from a total of approximately 408,000). Because many stock-days do not contain block trades we also estimate a simpler version of the VAR in which we pool dark and block trades into a single volume category. This allows greater coverage across the sample (approximately 223,000 stock-days). We also estimate a version of VAR in which we sign trades as buyer/seller-initiated using only

information that is readily available to market participants: trades with price above (below) the prevailing midquote are classified as buyer- (seller-) initiated and trades at the midquote are discarded. Our main results are robust to these alternative specifications.

10. Conclusions

Our analysis shows that trades executed in the dark tend to be less informed than trades executed in the lit market. Therefore, by disproportionately reducing the number of uninformed trades in the lit market, high levels of dark trading increase adverse selection risk and the lit market's bid-ask spreads, consistent with the theoretical predictions of Zhu (2014). The increased adverse selection risk and trading costs in the lit market increase the incentives for order flow to migrate away from the lit market, potentially leading to a self-reinforcing spiral.

As dark trading increases and adverse selection risks in the lit market become more severe, order book quotes take on a more important role in impounding new information compared to trade prices. Liquidity providers in the lit market become increasingly informed. This finding is consistent with the notion that informed traders have a comparative advantage in liquidity provision when adverse selection risks are high, due to their informational advantage. We also find that dark trades play a greater role in price discovery as the level of dark trading increases. However, the dark market's share of price discovery increases at a slower rate than its increase in market share, providing further evidence that dark trades tend to be less informed than lit trades. Informed traders in the lit market naturally would like to trade with the less-informed order flow in the dark, but their ability to do so is limited by lower execution probability in the dark due to their tendency to cluster on one side of the market (Zhu, 2014) and due to exclusivity of some dark pools that limit participation to relatively uninformed clientele (Boni et al., 2012).

The overall impact of these changes on informational efficiency depends on the level of dark trading. We show that for the typical stock, the level of dark trading is not harmful. However, for stock-days with high levels of dark trading, price discovery is harmed. Together, the results provide support for the concerns of regulators that high levels of dark trading harm informational efficiency and price discovery. High levels of

dark trading are found to be harmful throughout the sample period, in large and in small stocks. This does not, however, mean that dark trading in general or even in equilibrium is harmful. Our results indicate that low levels of dark trading do not have a negative impact on informational efficiency and might even be beneficial. For most stocks in our sample, the level of dark trading on a typical day is below harmful levels. This result has important policy implications. It provides evidence that regulatory action should consider the level of dark trading in specific stocks, rather than the aggregate market level of dark trading. Regulatory changes proposed in MiFID II to cap dark trading at 4% per venue and 8% across all venues in Europe could have unintended consequences, particularly if these caps are below harmful levels for the European market.

We find no evidence that block trades negotiated without pre-trade transparency harm informational efficiency. Block trades differ from dark trades in that upstairs brokers' unique role as 'information repositories' allows block trades to tap into additional liquidity that would not otherwise be expressed in the limit order book, thereby expanding aggregate liquidity (Grossman, 1992; Bessembinder and Venkataraman, 2004). Furthermore, block trades are largely uninformed, but due to their size they would cause significant temporary price distortions if submitted to the limit order book. In the upstairs market, a block broker can reduce adverse selection risk for the trade's counterparty by signaling the motivation for the trade, thereby reducing price impact and avoiding the temporary price distortions that would occur in the limit order book (Bessembinder and Venkataraman, 2004). Again, this result has important policy implications, providing evidence that regulation of dark trading needs to be carefully designed to account for the fact that not all dark trading has the same effects on price discovery. This result supports the decision by the European Commission to exclude trades executed using the 'Large in Scale' waiver from the 4% and 8% dark caps.

Appendix A. Informational efficiency measures

Our informational efficiency metrics follow the existing empirical literature. They measure the extent to which prices deviate from a random walk and/or are predictable using past information. In a perfectly informationally efficient, frictionless market, prices at all times equal the fundamental value (the expected value of the stock given all available information). Prices change only due to the arrival of new information. Because new information is unpredictable by definition, prices follow a martingale. Thus, price changes should not be predictable using past information (such as lagged stock or market returns), returns should have zero autocorrelation, and (with the additional assumption that innovations in the log fundamental are independent and identically distributed (i.i.d.) Gaussian) the variance of returns should increase linearly with the return horizon.

Many different imperfections and sources of inefficiency in real markets cause deviations from the characteristics of perfectly efficient markets. Trade prices have negative autocorrelation due to the mechanical effect of bid-ask bounce (e.g., Roll, 1984). This liquidity-related effect is not, however, present in midquotes. Midquotes can (and do) deviate from the characteristics expected under perfectly efficient markets, and importantly, all of the mechanisms that cause such deviations imply some degree of informational inefficiency. For example, inventory control models such as Ho and Stoll (1981) show that inventory management by liquidity providers causes negative autocorrelation in midquote returns. Risk-averse liquidity providers that hold positive inventory will adjust quotes upward so that the midquote is greater than the fundamental value to attract sellers and reduce their inventory position. The opposite occurs when liquidity providers are short. The negative autocorrelation in midquotes (and ability to predict returns using past order flow) is associated with midquote deviations from fundamental values, i.e., informational inefficiency. As another example, consider price formation models such as the sequential trade model of Glosten and Milgrom (1985) and the Kyle (1985) model with repeated auctions. In both models, prices are initially efficient with respect to public information and gradually approach strong-form efficiency as private information is revealed through the course of trading. Convergence of prices toward the full-information values (expected fundamental value given all public *and* private information) causes positive short-run autocorrelation in expost midquote

returns. Importantly, the autocorrelation occurs as a result of prices being less that fully informationally efficient; once prices reflect all public and private information, returns no longer display autocorrelation, and if private information were revealed instantaneously rather than gradually, returns would also have zero autocorrelation. Finally, both under- and overreaction to information, as well as delayed reaction to information (e.g., non-synchronous trading, stale prices) cause midquotes to deviate from the characteristics expected under perfect efficiency. In summary, the many different mechanisms that cause midquotes to deviate from the characteristics expected under a perfectly efficient market imply some degree of informational efficiency.

Both positive and negative midquote autocorrelations imply less-than-perfect informational efficiency. We calculate first-order return autocorrelations for each stock-day, at various intraday frequencies, $k \in \{10\sec, \ 30\sec, \ 60\sec\}$, similar to Hendershott and Jones (2005):

$$Autocorrelation_k = Corr(r_{k,t}, r_{k,t-1}), \tag{A.1}$$

where $r_{k,t}$ is the tth midquote return of length k for a stock-day (stock-day subscripts are suppressed). Taking the absolute value of the autocorrelation gives a measure of informational efficiency that captures both under- and overreaction of returns to information, with larger values indicating greater inefficiency. We compute a combined autocorrelation measure, $Autocorrelation_{Factor}$, by taking the first principal component of the absolute autocorrelations at the three frequencies, and then scaling the measure so that it ranges from zero (highly efficient) to 100 (highly inefficient). Using the first principal component is a way of summarizing the results across informational efficiency metrics calculated at different frequencies and can help reduce error in the individual proxies.[34]

If a stock's price follows a random walk, the variance of its returns is a linear function of the measurement frequency, i.e., $\sigma^2_{k-periodReturn}$ is k times larger than $\sigma^2_{1-periodReturn}$. The variance ratio exploits this property to measure inefficiency as a price series' deviation from the characteristics that would be expected under a random

---

[34] We expect the informational efficiency metrics calculated at different frequencies to correlate with the underlying latent variable (informational inefficiency) but each will also contain some measurement error. Therefore, the metrics at different frequencies will have some common variance arising from the variance in informational inefficiency. The first principal component is the linear combination of the different frequency metrics that explains the maximal amount of common variance and thus should be closely related to the underlying latent variable, while containing less noise if the measurement errors are less than perfectly correlated across the different frequencies.

walk (e.g., Lo and MacKinlay, 1988). We calculate three variance ratios for each stock-day at different intraday frequencies:

$$VarianceRatio_{kl} = \left| \frac{\sigma_{kl}^2}{k\sigma_l^2} - 1 \right|, \tag{A.2}$$

where σ_l^2 and σ_{kl}^2 are the variances of l-second and kl-second midquote returns for a given stock-day. We use the (l, kl) combinations: (1-sec, 10-sec), (10-sec, 60-sec), (1-min, 5-min). We compute a combined variance ratio, $VarianceRatio_{Factor}$, by taking the first principal component of the three variance ratios, and then scaling the measure so that it ranges from zero (highly efficient) to 100 (highly inefficient).

Our third measure of informational efficiency is an intraday adaptation of the Hou and Moskowitz (2005) *Delay*, i.e., the extent to which lagged market returns predict a stock's midquote returns. For each stock-day we estimate a regression of one-minute midquote returns for stock i, $r_{i,t}$, on the All Ordinaries market index return, $r_{m,t}$, and ten lags (suppressing day subscripts):

$$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \sum_{k=1}^{10} \delta_{i,k} r_{m,t-k} + \varepsilon_{it} . \tag{A.3}$$

We save the R^2 from the above unconstrained regression, $R^2_{Unconstrained}$, and re-estimate the regression constraining the coefficients on lagged market returns to zero, $\delta_{i,k} = 0, \forall k$, again saving the R^2, $R^2_{Constrained}$. *Delay* is then calculated as:

$$Delay = 100 \left(1 - \frac{R^2_{Constrained}}{R^2_{Unconstrained}} \right), \tag{A.4}$$

and takes values between zero and 100. The larger this measure, the more variation in stock returns is explained by lagged market returns, which implies more sluggish incorporation of market-wide information into the stock's price, and, therefore, lower informational efficiency.

References

Admati, A.R., Pfleiderer, P., 1988. A theory of intraday patterns: Volume and price variability. Review of Financial Studies 1, 3–40.

Anand, A., Subrahmanyam, A., 2008. Information and the intermediary: Are market intermediaries informed traders in electronic markets? Journal of Financial and Quantitative Analysis 43, 1–28.

Anderson, R.M., Eom, K.S., Hahn, S.B., Park, J.-H., 2013. Autocorrelation and partial price adjustment. Journal of Empirical Finance 24, 78–93.

Australian Securities and Investments Commission, 2013. Report 331, Dark liquidity and high frequency trading.

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

Baruch, S., 2005. Who benefits from an open limit order book? Journal of Business 78, 1267–1306.

Bessembinder, H., Venkataraman, K., 2004. Does an electronic stock exchange need an upstairs market? Journal of Financial Economics 73, 3–36.

Black, F., 1986. Noise. Journal of Finance 41, 529–543.

Bloomfield, R., O'Hara, M., Saar, G., 2005. The "make or take" decision in an electronic market: Evidence on the evolution of liquidity. Journal of Financial Economics 75, 165–199.

Boehmer, E., Saar, G., Yu, L., 2005. Lifting the veil: An analysis of pre-trade transparency at the NYSE. Journal of Finance 60, 783–815.

Boni, L., Brown, D.C., Leach, J.C., 2012. Dark pool exclusivity matters. Unpublished working paper, University of New Mexico, University of Arizona, and University of Colorado at Boulder.

Booth, G.G., Lin, J., Martikainen, T., Tse, Y., 2002. Trading and pricing in upstairs and downstairs stock markets. Review of Financial Studies 15, 1111–1135.

Boulatov, A., George, T.J., 2013. Hidden and displayed liquidity in securities markets with informed liquidity providers. Review of Financial Studies 26, 2095–2137.

Bound, J., Jaeger, D.A., Baker, R.M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous

explanatory variable is weak. Journal of the American Statistical Association 90, 443–450.

Buti, S., Rindi, B., Werner, I.M., 2011. Diving into dark pools. Unpublished working paper, University of Toronto, Bocconi University, and The Ohio State University.

Buti, S., Rindi, B., Werner, I.M., 2013. Dark pool trading strategies, market quality and welfare, Unpublished working paper, University of Toronto, Bocconi University, and The Ohio State University.

Chordia, T., Roll, R., Subrahmanyam, A., 2008. Liquidity and market efficiency. Journal of Financial Economics 87, 249–268.

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

Colliard, J.-E., Foucault, T., 2012. Trading fees and efficiency in limit order markets. Review of Financial Studies 25, 3389–3421.

Degryse, H., de Jong, F., van Kervel, V., 2014. The impact of dark trading and visible fragmentation on market quality. Review of Finance, forthcoming.

Eom, K.S., Ok, J., Park, J., 2007. Pre-trade transparency and market quality. Journal of Financial Markets 10, 319–341.

Foucault, T., Menkveld, A.J., 2008. Competition for order flow and smart order routing systems. Journal of Finance 63, 19–58.

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.

Goettler, R.L., Parlour, C.A., Rajan, U., 2009. Informed traders and limit order markets. Journal of Financial Economics 93, 67–87.

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

Grossman, S., 1992. The informational role of upstairs and downstairs markets. Journal of Business 65, 509–529.

Grossman, S., Stiglitz, J., 1980. On the impossibility of informationally efficient markets. American Economic Review 10, 393–408.

Hasbrouck, J., 1991. Measuring the information content of stock trades. Journal of Finance 46, 179–207.

Hasbrouck, J., 1995. One security, many markets: Determining the contributions to price discovery. Journal of Finance 50, 1175–1199.

Hasbrouck, J., Saar, G., 2013. Low-latency trading. Journal of Financial Markets 16, 646–679.

Hautsch, N., Huang, R., 2012. The market impact of a limit order. Journal of Economic Dynamics and Control 36, 501–522.

Hendershott, T., Jones, C.M., 2005. Island goes dark: Transparency, fragmentation, and regulation. Review of Financial Studies 18, 743–793.

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

Hendershott, T., Mendelson, H., 2000. Crossing networks and dealer markets: Competition and performance. Journal of Finance 55, 2071–2115.

Ho, T.S.Y., Stoll, H.R., 1981. Optimal dealer pricing under transactions and return uncertainty. Journal of Financial Economics 9, 47–73.

Hou, K., Moskowitz, T.J., 2005. Market frictions, price delay, and the cross-section of expected returns. Review of Financial Studies 18, 981–1020.

Kaniel, R., Liu, H., 2006. So what orders do informed traders use? Journal of Business 79, 1867–1913.

Kyle, A.S., 1981. An equilibrium model of speculation and hedging. Doctoral dissertation. University of Chicago.

Kyle, A.S., 1984. Market structure, information, futures markets, and price formation. In: Storey, G.G., Schmitz, A., Sarris, A.H. (Eds.), International Agricultural Trade: Advanced Readings in Price Formation, Market Structure, and Price Instability. Westview Press, Boulder and London, pp. 45–64.

Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica 53, 1315–1335.

Kyle, A.S., 1989. Informed speculation with imperfect competition. Review of Economic Studies 56, 317–356.

Kwan, A., Masulis, R., McInish, T.H., 2015. Trading rules, competition for order flow and market fragmentation. Journal of Financial Economics 115, 330–348.

Lo, A., MacKinlay, C., 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. Review of Financial Studies 1, 41–66.

Madhavan, A., 1995. Consolidation, fragmentation, and the disclosure of trading information. Review of Financial Studies 8, 579–603.

Madhavan, A., Cheng, M., 1997. In search of liquidity: Block trades in the upstairs and downstairs markets. Review of Financial Studies 10, 175–203.

Madhavan, A., Porter, D., Weaver, D., 2005. Should securities markets be transparent? Journal of Financial Markets 8, 266–288.

Nimalendran, M., Ray, S., 2014. Informational linkages between dark and lit trading venues. Journal of Financial Markets 17, 230–261.

O'Hara, M., Ye, M., 2011. Is market fragmentation harming market quality? Journal of Financial Economics 100, 459–474.

Pagano, M., Röell, A.A., 1996. Transparency and liquidity: A comparison of auction and dealer markets with informed trading. Journal of Finance 51, 579–611.

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

Ready, M.J., 2013. Determinants of volume in dark pool crossing networks. Unpublished working paper, University of Wisconsin-Madison.

Rindi, B., 2008. Informed traders as liquidity providers: Anonymity, liquidity and price formation. Review of Finance 12, 497–532.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. Journal of Finance 39, 1127–1139.

Rösch, D.M., Subrahmanyam, A., van Dijk, M.A., 2013. An empirical analysis of co-movements in high- and low-frequency measures of market efficiency. Unpublished working paper, Rotterdam School of Management, and University of California, Los Angeles.

Roşu, I., 2013. Liquidity and information in order driven markets. Unpublished working paper, HEC Paris.

Securities and Exchange Commission (SEC), 2010. Concept Release on Equity Market Structure, Release # 34-61358, File # S7-02-10.

Seppi, D., 1990. Equilibrium block trading and asymmetric information. Journal of Finance 45, 73–94.

Tuttle, L., 2013. Alternative trading systems: Description of ATS trading in national market system stocks. Division of Economic and Risk Analysis, Securities and Exchange Commission.

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

Ye, M., 2012. Price manipulation, price discovery and transaction costs in the crossing network. Unpublished working paper, University of Illinois at Urbana-Champaign.

Yin, X., 2005. A comparison of centralized and fragmented markets with costly search. Journal of Finance 60, 1567–1590.

Zhu, H., 2014. Do dark pools harm price discovery? Review of Financial Studies 27, 747–789.

Table 1
Descriptive statistics
This table reports means, standard deviations, and quartile points (*P25*, *Median*, *P75*) of variables calculated at the stock-day level. Total volume consists of *Lit* trades (trades executed in the transparent central limit order book), *Dark* trades (trades executed without pre-trade transparency below block size), and *Block* trades (large trades executed without pre-trade transparency). *Price* is the closing bid-ask midquote. *Constrained spread* measures the proportion of trading day for which the stock's spread is constrained to one tick size. *Volatility* is the standard deviation of one-minute midquote returns. *Message-to-trade* is the ratio of number of order messages (including order entry, amendment, and cancellation) to the number of trades. The sample comprises the 500 largest ASX-listed stocks from February 1, 2008 to October 30, 2011.

	Mean	Std. dev.	P25	Median	P75
Volumes and trades					
Total volume ($ mil)	9.91	38.75	0.12	0.70	4.49
Total trades (count)	1,050	1,959	41	268	1,267
Stock characteristics					
Market capitalization ($ mil)	2,749	9,482	193	422	1,553
Price ($)	4.76	9.77	0.73	1.79	4.42
Quoted spread (bps)	128.51	172.24	32.31	66.54	158.07
Constrained spread (fraction)	0.60	0.36	0.29	0.71	0.93
Volatility (bps)	16.61	14.64	8.01	12.77	20.60
Message-to-trade (ratio)	4.58	34.48	2.56	3.56	4.90

45

Table 2
Correlations between independent variables

This table reports correlations between independent variables. *DARK* and *BLOCK* are the percentage of the stock-day's total dollar volume executed without pre-trade transparency below block size and at block size, respectively. *Market capitalization*, *Quoted spread* (time-weighted average of the stock-day's limit order book proportional quoted spread), and *Total $ volume* (comprising dark, block, and lit limit order book volume) are in logs. *Constrained spread* measures the proportion of trading day for which the stock's spread is constrained to one tick size. *Volatility* is the standard deviation of one-minute midquote returns. *Message-to-trade* is the ratio of number of quote messages (including order entry, amendment, and cancellation) to the number of trades. The sample comprises the 500 largest ASX-listed stocks from February 1, 2008 to October 30, 2011.

	DARK	BLOCK	Market capitalization	Total $ volume	Quoted spread	Volatility	Constrained spread	Message-to-trade
DARK	1	0.01	0.13	0.23	-0.10	-0.06	0.11	0.01
BLOCK	0.01	1	0.24	0.29	-0.21	-0.06	0.09	0.00
Market capitalization	0.13	0.24	1	0.79	-0.77	-0.30	0.27	0.01
Total $ volume	0.23	0.29	0.79	1	-0.77	-0.11	0.52	-0.01
Quoted spread	-0.10	-0.21	-0.77	-0.77	1	0.40	-0.40	0.00
Volatility	-0.06	-0.06	-0.30	-0.11	0.40	1	-0.11	-0.01
Constrained spread	0.11	0.09	0.27	0.52	-0.40	-0.11	1	-0.02
Message-to-trade	0.01	0.00	0.01	-0.01	0.00	-0.01	-0.02	1

Table 3

Informativeness of trade types

This table reports means, standard deviations, and quartile points (*P25*, *Median*, *P75*) of trade informativeness variables calculated at the stock-day level. $PriceImpact_{LIT}$, $PriceImpact_{DARK}$, and $PriceImpact_{BLOCK}$ are the permanent price impacts (bps/\$10,000) of lit, dark, and block volume calculated from the cumulative impulse response functions from a vector auto-regression model. *Lit* trades are trades executed in the transparent central limit order book, *Dark* trades are trades executed without pre-trade transparency below block size, and *Block* trades are large trades executed without pre-trade transparency. The sample comprises the 500 largest ASX-listed stocks from February 1, 2008 to October 30, 2011.

	Mean (equal-weighted)	Mean (\$volume-weighted)	Std. dev.	P25	Median	P75
$PriceImpact_{LIT}$	3.62	0.65	4.94	0.69	1.91	4.74
$PriceImpact_{DARK}$	3.31	0.41	27.94	-0.10	0.03	1.28
$PriceImpact_{BLOCK}$	0.15	0.02	2.53	-0.02	0.01	0.11

Table 4

Effects of dark and block trading on the bid-ask spread

   This table reports regression estimates using a stock-day panel, in which the dependent variable is the log time-weighted average proportional quoted bid-ask spread in the central limit order book. The key independent variables, *DARK* and *BLOCK*, are the percentage of the stock-day's total dollar volume executed without pre-trade transparency below block size and at block size, respectively. We report five models: (i) one-stage OLS, (ii) one-stage OLS with stock fixed effects; (iii) one-stage OLS with date fixed effects; (iv) two-stage least squares (2SLS) using the first set of instruments (a dummy variable for the removal of the ten-second rule that restricted dark trading, a dummy variable for a change in exchange fees and the introduction of Centre Point, and the number of dark pools in operation, as well as its square); and (v) 2SLS using the second set of instruments (the average of *DARK* and *BLOCK* on the same day for all other stocks in the relevant size quartile). In the first stage of the 2SLS models we regress *DARK* and *BLOCK* on the instrumental variables and control variables, and in the second stage we regress each of the dependent variables on fitted values of *DARK* and *BLOCK* from the first-stage regressions, and control variables. *Market capitalization* and *Total $ volume* (comprising dark, block, and lit limit order book volume) are in logs. *Constrained spread* measures the proportion of trading day for which the stock's spread is constrained to one tick size. *Volatility* is the standard deviation of one-minute midquote returns. *Message-to-trade* is the ratio of number of quote messages (including order entry, amendment, and cancellation) to the number of trades. *Time* is a linear time trend starting at zero and incrementing by one for every date in our sample. $Y_{OtherStocks}$ is the average of the dependent variable (*Autocorrelation$_{Factor}$*, *VarianceRatio$_{Factor}$*, or *Delay*), on the same day for all other stocks in the relevant size quartile. R^2 estimates exclude the variance explained by the fixed effects. Standard errors are clustered both by stock and by date and *t*-statistics are reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, And 10% levels. The sample comprises the 500 largest ASX-listed stocks from February 1, 2008 to October 30, 2011.

	Log quoted spread				
Intercept	8.492	-4.254	-4.273	8.651	7.589
	(77.37)***	(-102.4)***	(-211.1)***	(72.91)***	(25.51)***
DARK	0.006	0.002	0.006	0.019	0.021
	(11.05)***	(8.77)***	(11.00)***	(10.47)***	(10.49)***
BLOCK	0.004	0.002	0.004	-0.005	-0.005
	(7.52)***	(2.99)***	(7.84)***	(-1.10)	(-1.82)*
Market capitalization	-0.216	-0.518	-0.210	-0.189	-0.130
	(-6.56)***	(-16.12)***	(-6.54)***	(-5.66)***	(-3.51)***
Constrained spread	-0.117	-1.022	-0.040	-0.066	-0.057
	(-1.89)*	(-24.08)***	(-0.71)	(-1.11)	(-0.91)
Total $ volume	-0.233	-0.093	-0.241	-0.261	-0.260
	(-12.68)***	(-8.92)***	(-13.53)***	(-12.57)***	(-12.70)***
Volatility	0.018	0.008	0.017	0.019	0.019
	(5.52)***	(4.51)***	(4.89)***	(5.23)***	(5.38)***
Message-to-trade	0.000	0.000	0.000	0.000	0.000
	(-0.09)	(2.23)**	(0.08)	(-0.87)	(-1.10)
Time				0.000	
				(-3.69)***	
$Y_{OtherStocks}$					0.136
					(3.18)***
R^2	0.74	0.15	0.68	0.75	0.75
Estimation method	OLS	OLS	OLS	2SLS	2SLS
Fixed effects	None	Stock	Date	None	None
Instrumental variables	None	None	None	Set 1	Set 2

Table 5

Effects of dark and block trading on price discovery shares

This table reports regression estimates using a stock-day panel, in which the dependent variables are the information leadership share of lit trade prices (trades executed in the transparent central limit order book) relative to dark and block trade prices (ILS_{LIT}), and the information leadership share of midquotes relative to trade prices ($ILS_{MIDQUOTE}$). Both ILS_{LIT} and $ILS_{MIDQUOTE}$ are scaled up by a factor of 100. The key independent variables, *DARK* and *BLOCK*, are the percentage of the stock-day's total dollar volume executed without pre-trade transparency below block size and at block size, respectively. For each dependent variable we report five models: (i) one-stage OLS, (ii) one-stage OLS with stock fixed effects; (iii) one-stage OLS with date fixed effects; (iv) two-stage least squares (2SLS) using the first set of instruments (a dummy variable for the removal of the ten-second rule that restricted dark trading, a dummy variable for a change in exchange fees and the introduction of Centre Point, and the number of dark pools in operation, as well as its square); and (v) 2SLS using the second set of instruments (the average of *DARK* and *BLOCK* on the same day for all other stocks in the relevant size quartile. In the first stage of the 2SLS models we regress *DARK* and *BLOCK* on the instrumental variables and control variables, and in the second stage we regress each of the dependent variables on fitted values of *DARK* and *BLOCK* from the first-stage regressions, and control variables. *Market capitalization*, *Quoted spread* (time-weighted average of the stock-day's limit order book proportional quoted spread), and *Total $ volume* (comprising dark, block, and lit limit order book volume) are in logs. *Constrained spread* measures the proportion of trading day for which the stock's spread is constrained to one tick size. *Volatility* is the standard deviation of one-minute midquote returns. *Time* is a linear time trend starting at zero and incrementing by one for every date in our sample. *Message-to-trade* is the ratio of number of quote messages (including order entry, amendment, and cancellation) to the number of trades. $Y_{OtherStocks}$ is the average of the dependent variable ($Autocorrelation_{Factor}$, $VarianceRatio_{Factor}$, or *Delay*), on the same day for all other stocks in the relevant size quartile. R^2 estimates exclude the variance explained by the fixed effects. Standard errors are clustered both by stock and by date and *t*-statistics are reported in parentheses. ***, **, And * indicate statistical significance at 1%, 5%, and 10% levels. The sample comprises the 500 largest ASX-listed stocks from February 1, 2008 to October 30, 2011.

	ILS_{LIT}					$ILS_{MIDQUOTE}$				
Intercept	36.177	-0.803	-7.101	29.953	14.468	66.321	-0.099	0.092	79.291	37.347
	(11.57)***	(-5.65)***	(-37.52)***	(8.29)***	(4.09)***	(21.21)***	(-1.31)	(0.57)	(24.35)***	(8.07)***
DARK	-0.004	-0.021	-0.015	-0.023	-0.024	0.085	0.029	0.090	0.311	0.321
	(-0.51)	(-3.35)***	(-2.18)**	(-0.99)	(-0.99)	(13.35)***	(5.53)***	(15.11)***	(13.83)***	(12.78)***
BLOCK	-0.129	-0.126	-0.090	-0.131	-0.287	-0.013	-0.035	-0.007	-0.285	-0.220
	(-15.39)***	(-14.80)***	(-11.55)***	(-2.99)***	(-9.65)***	(-1.48)	(-4.47)***	(-0.83)	(-5.56)***	(-6.21)***
Market capitalization	0.510	0.830	0.675	0.643	-0.652	-0.692	-0.533	-0.742	-0.479	0.416
	(1.84)*	(1.51)	(2.33)**	(2.16)**	(-2.31)**	(-2.60)***	(-1.34)	(-2.84)***	(-1.95)*	(1.64)
Quoted spread	-3.810	-3.020	-3.537	-3.489	-3.417	0.444	0.458	0.217	-0.545	-0.548
	(-10.11)***	(-3.35)**	(-9.42)***	(-8.37)***	(-9.38)***	(1.04)	(0.91)	(0.58)	(-1.29)	(-1.20)
Constrained spread	9.017	5.513	8.203	7.711	7.511	-3.689	-4.849	-2.155	-2.090	-2.114
	(12.31)***	(5.16)***	(12.50)***	(11.16)***	(10.46)***	(-6.20)***	(-9.12)***	(-3.55)***	(-3.70)***	(-3.65)***
Total $ volume	2.767	2.057	2.624	2.941	2.940	-0.643	0.603	-0.826	-1.239	-1.252
	(14.25)***	(11.41)***	(13.86)***	(12.56)***	(14.48)***	(-3.25)***	(3.03)***	(-4.67)***	(-5.31)***	(-5.01)***
Volatility	0.059	0.044	0.057	0.072	0.057	0.137	0.140	0.119	0.134	0.147
	(2.27)**	(2.11)**	(2.11)**	(2.21)**	(2.25)**	(4.53)***	(4.11)***	(4.16)***	(4.26)***	(4.28)***
Message-to-trade	0.001	0.022	-0.043	-0.054	-0.030	0.001	0.001	0.002	0.002	0.001
	(0.04)	(1.19)	(-2.28)**	(-2.58)***	(-1.56)	(0.65)	(0.66)	(0.99)	(0.83)	(0.66)
Time			0.005					-0.009		
			(5.53)***					(-15.19)***		
$Y_{OtherStocks}$				0.376					0.571	
				(10.08)***					(19.23)***	
R^2	0.11	0.02	0.11	0.12	0.13	0.04	0.02	0.03	0.05	0.06
Estimation method	OLS	OLS	OLS	2SLS	2SLS	OLS	OLS	OLS	2SLS	2SLS
Fixed effects	None	Stock	Date	None	None	None	Stock	Date	None	None
Instrumental variables	None	None	None	Set 1	Set 2	None	None	None	Set 1	Set 2

Table 6

OLS regressions for high-frequency measures of informational efficiency

This table reports OLS regression estimates using a stock-day panel, in which the dependent variables are high-frequency estimates of informational inefficiency, which range from zero (perfect efficiency) to 100 (complete inefficiency). $Autocorrelation_{Factor}$ and $VarianceRatio_{Factor}$ are the first principal components of absolute autocorrelations of midquote returns and variance ratios at different intraday frequencies. *Delay* measures intraday midquote return predictability using lagged market returns. *DARK* and *BLOCK* are the percentage of the stock-day's total dollar volume executed without pre-trade transparency below block size and at block size, respectively. *Market capitalization*, *Quoted spread* (time-weighted average of the stock-day's limit order book proportional quoted spread), and *Total $ volume* (comprising dark, block, and lit limit order book volume) are in logs. *Constrained spread* measures the proportion of trading day for which the stock's spread is constrained to one tick size. *Volatility* is the standard deviation of one-minute midquote returns. *Message-to-trade* is the ratio of number of quote messages (including order entry, amendment, and cancellation) to the number of trades. The regression model is estimated for each dependent variable without fixed effects, with stock fixed effects, and with date fixed effects. R^2 estimates exclude the variance explained by the fixed effects. Standard errors are clustered both by stock and by date and *t*-statistics are reported in parentheses. ***, **, And * indicate statistical significance at 1%, 5%, and 10% levels. The sample comprises the 500 largest ASX-listed stocks from February 1, 2008 to October 30, 2011.

	$Autocorrelation_{Factor}$			$VarianceRatio_{Factor}$			Delay		
Intercept	0.251	-0.046	-0.091	1.980	-0.024	-0.068	92.905	0.029	0.171
	(0.22)	(-2.33)**	(-1.18)	(2.49)**	(-1.35)	(-1.09)	(36.02)***	(0.47)	(1.04)
DARK	0.042	0.018	0.041	0.029	0.012	0.030	0.048	0.004	0.048
	(16.84)***	(8.63)***	(16.57)***	(17.75)***	(8.07)***	(17.67)***	(8.72)***	(1.83)*	(8.95)***
BLOCK	-0.013	-0.018	-0.015	-0.006	-0.011	-0.009	-0.002	-0.001	-0.006
	(-5.15)***	(-8.53)***	(-6.19)***	(-3.55)***	(-7.50)***	(-5.13)***	(-0.26)	(-0.27)	(-1.02)
Market capitalization	-0.334	-0.571	-0.333	-0.328	-0.396	-0.321	-1.923	-0.633	-1.830
	(-2.74)***	(-4.79)***	(-2.70)***	(-3.82)***	(-4.54)***	(-3.71)***	(-6.79)***	(-3.85)***	(-6.55)***
Quoted spread	-0.120	-0.542	-0.130	-0.290	-0.391	-0.326	3.028	1.697	3.267
	(-0.90)	(-4.18)***	(-0.99)	(-3.22)***	(-5.51)***	(-3.66)***	(10.19)***	(8.64)***	(11.79)***
Constrained spread	-0.553	-2.267	-0.584	-0.397	-1.950	-0.316	7.442	4.597	6.563
	(-2.28)**	(-12.82)***	(-2.32)**	(-2.05)**	(-14.55)***	(-1.55)	(12.71)***	(9.82)***	(11.43)***
Total $ volume	0.817	1.144	0.826	0.661	0.880	0.647	-0.754	0.115	-0.674
	(13.80)***	(23.45)***	(14.10)***	(16.39)***	(37.78)***	(15.91)***	(-5.98)***	(2.22)**	(-6.08)***
Volatility	0.023	0.027	0.023	-0.002	-0.003	-0.004	-0.067	-0.034	-0.041
	(3.75)***	(3.53)***	(3.43)***	(-0.95)	(-1.94)*	(-2.20)**	(-4.91)***	(-5.53)***	(-4.64)***
Message-to-trade	0.004	0.004	0.004	0.004	0.003	0.004	0.000	0.000	0.001
	(1.59)	(1.52)	(1.60)	(1.60)	(1.51)	(1.61)	(1.31)	(0.41)	(2.17)**
R^2	0.06	0.04	0.06	0.10	0.04	0.10	0.17	0.01	0.18
Estimation method	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Fixed effects	None	Stock	Date	None	Stock	Date	None	Stock	Date

50

Table 7

Instrumental variables regressions for high-frequency measures of informational efficiency

This table reports estimates from a two-stage least squares regression (2SLS), using two different sets of instrumental variables. *DARK* and *BLOCK* are the percentage of the stock-day's total dollar volume executed without pre-trade transparency below block size and at block size, respectively. In the first stage *DARK* and *BLOCK* are regressed on the instrumental variables and control variables. The first set of instrumental variables comprises a dummy variable for the removal of the ten-second rule that restricted dark trading ($D_t^{NO\_10SEC\_RULE}$), a dummy variable for a change in exchange fees and the introduction of Centre Point ($D_t^{NEW\_FEES}$), and the number of dark pools in operation, *DarkVenues_t* as well as its square, *DarkVenues*$_t^2$. The second set of instrumental variables ($DARK_i^{NOT}$ and $BLOCK_i^{NOT}$) is the average of *DARK* and *BLOCK*, respectively, on the same day for all other stocks in the relevant size (market capitalization) quartile. In the second stage we regress each of the dependent variables on fitted values of *DARK* and *BLOCK* from the first-stage regressions, and control variables. The dependent variables are high-frequency estimates of informational inefficiency, which range from zero (perfect efficiency) to 100 (complete inefficiency). *Autocorrelation_Factor* and *VarianceRatio_Factor* are the first principal components of absolute autocorrelations of midquote returns and variance ratios at different intraday frequencies. *Delay* measures intraday midquote return predictability using lagged market returns. *Market capitalization*, *Quoted spread* (time-weighted average of the stock-day's limit order book proportional quoted spread), and *Total $ volume* (comprising dark, block, and lit limit order book volume) are in logs. *Constrained spread* measures the proportion of trading day for which the stock's spread is constrained to one tick size. *Volatility* is the standard deviation of one-minute midquote returns. *Message-to-trade* is the ratio of number of quote messages (including order entry, amendment, and cancellation) to the number of trades. *Time* is a linear time trend starting at zero and incrementing by one for every date in our sample. *Y_OtherStocks* is the average of the dependent variable (*Autocorrelation_Factor*, *VarianceRatio_Factor* or *Delay*), on the same day for all other stocks in the relevant size quartile. Standard errors are clustered both by stock and by date and *t*-statistics are reported in parentheses. ***, **, And * indicate statistical significance at 1%, 5%, and 10% levels. The sample comprises the 500 largest ASX-listed stocks from February 1, 2008 to October 30, 2011.

|	Autocorrelation_Factor		VarianceRatio_Factor		Delay	
Intercept	3.052 (2.41)**	0.803 (0.62)	4.984 (6.06)***	1.949 (2.30)**	97.233 (37.57)***	49.310 (8.12)***
DARK	0.142 (15.91)***	0.136 (16.48)***	0.106 (16.47)***	0.099 (18.28)***	0.247 (10.75)***	0.262 (10.04)***
BLOCK	-0.104 (-5.17)***	-0.083 (-6.61)***	-0.045 (-2.96)***	-0.043 (-4.63)***	-0.273 (-5.14)***	-0.210 (-5.47)***
Market capitalization	-0.215 (-1.87)*	-0.534 (-4.15)***	-0.243 (-3.05)***	-0.616 (-6.65)***	-1.680 (-6.82)***	-0.515 (-1.71)*
Quoted spread	-0.463 (-3.05)***	-0.430 (-2.82)***	-0.581 (-6.49)***	-0.500 (-5.28)***	2.393 (8.92)***	2.476 (9.27)***
Constrained spread	-0.576 (-2.52)**	-0.649 (-2.94)***	-0.270 (-1.48)	-0.538 (-3.12)***	7.279 (12.95)***	7.086 (13.26)***
Total $ volume	0.596 (7.55)***	0.582 (7.52)***	0.446 (9.20)***	0.466 (9.71)***	-1.087 (-7.63)***	-1.072 (-7.36)***
Volatility	0.037 (3.81)***	0.033 (3.77)***	0.008 (2.97)***	0.006 (2.50)**	-0.044 (-4.51)***	-0.029 (-4.36)***
Message-to-trade	0.004 (1.57)	0.004 (1.57)	0.003 (1.57)	0.003 (1.56)	0.000 (-0.39)	0.000 (-0.39)
Time	0.000 (1.05)		0.000 (-2.02)**		0.000 (0.47)	
Y_OtherStocks		0.501 (15.4)***		0.648 (18.56)***		0.456 (10.94)***
R^2	0.07	0.08	0.11	0.13	0.18	0.20
Estimation method	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Instrumental variables	Set 1	Set 2	Set 1	Set 2	Set 1	Set 2

Table 8

Instrumental variables regressions for low-frequency measures of informational efficiency

This table reports estimates from a two-stage least squares regression (2SLS), using two different sets of instrumental variables. *DARK* and *BLOCK* are the percentage of the stock-day's total dollar volume executed without pre-trade transparency below block size and at block size, respectively. In the first stage *DARK* and *BLOCK* are regressed on the instrumental variables and control variables. The first set of instrumental variables comprises a dummy variable for the removal of the ten-second rule that restricted dark trading ($D_t^{NO\_10SEC\_RULE}$), a dummy variable for a change in exchange fees and the introduction of Centre Point ($D_t^{NEW\_FEES}$), and the number of dark pools in operation, *DarkVenues_t* as well as its square, $DarkVenues_t^2$. The second set of instrumental variables ($DARK_i^{NOT}$ and $BLOCK_i^{NOT}$) is the average of *DARK* and *BLOCK*, respectively, on the same day for all other stocks in the relevant size (market capitalization) quartile. In the second stage we regress each of the dependent variables on fitted values of *DARK* and *BLOCK* from the first-stage regressions, and control variables. The dependent variables are low-frequency estimates of informational inefficiency, which range from zero (perfect efficiency) to 100 (complete inefficiency). *Autocorrelation_Factor* and *VarianceRatio_Factor* are the first principal components of absolute autocorrelations of midquote returns and variance ratios at different daily frequencies. *Delay* measures intraday midquote return predictability using lagged market returns. *Market capitalization, Quoted spread* (time-weighted average of the stock-day's limit order book proportional quoted spread), and *Total $ volume* (comprising dark, block, and lit limit order book volume) are in logs. *Constrained spread* measures the proportion of trading day for which the stock's spread is constrained to one tick size. *Volatility* is the standard deviation of one-minute midquote returns. *Message-to-trade* is the ratio of number of quote messages (including order entry, amendment, and cancellation) to the number of trades. *Time* is a linear time trend starting at zero and incrementing by one for every date in our sample. *Y_OtherStocks* is the average of the dependent variable (*Autocorrelation_Factor, VarianceRatio_Factor* or *Delay*), on the same day for all other stocks in the relevant size quartile. Standard errors are clustered both by stock and by date and *t*-statistics are reported in parentheses. ***, **, And * indicate statistical significance at 1%, 5%, and 10% levels. The sample comprises the 500 largest ASX-listed stocks from February 1, 2008 to October 30, 2011.

|	$Autocorrelation_{Factor}$		$VarianceRatio_{Factor}$		$Delay$	
Intercept	35.234	15.403	22.138	12.619	145.577	47.631
	(19.62)***	(6.50)***	(19.99)***	(7.47)***	(26.00)***	(7.41)***
DARK	0.026	0.030	0.016	0.019	0.501	0.570
	(1.69)*	(2.10)**	(1.76)*	(2.14)**	(12.70)***	(13.01)***
BLOCK	0.089	0.054	0.031	0.017	-0.063	-0.111
	(2.44)**	(2.33)**	(1.30)	(1.13)	(-0.71)	(-1.90)*
Market capitalization	1.004	1.007	0.307	0.412	-1.265	2.376
	(7.34)***	(7.37)***	(2.84)***	(3.72)***	(-2.72)***	(4.79)***
Quoted spread	0.295	0.271	0.417	0.380	-1.092	-0.600
	(1.59)	(1.45)	(3.51)***	(3.22)***	(-1.92)*	(-1.12)
Constrained spread	-0.147	-0.231	-0.614	-0.524	1.600	0.121
	(-0.44)	(-0.72)	(-2.68)***	(-2.33)**	(1.59)	(0.12)
Total $ volume	-0.808	-0.769	-0.360	-0.354	-4.313	-3.946
	(-8.45)***	(-8.28)***	(-5.28)***	(-5.37)***	(-16.12)***	(-15.72)***
Volatility	-0.009	-0.009	0.014	0.010	-0.038	0.043
	(-1.82)*	(-1.82)*	(2.75)***	(2.37)**	(-2.50)**	(3.36)***
Message-to-trade	0.000	0.000	0.000	0.000	-0.005	-0.005
	(0.35)	(0.32)	(0.19)	(0.10)	(-1.71)*	(-1.76)*
Time	0.000		0.000		-0.012	
	(0.14)		(-0.98)		(-9.70)***	
$Y_{OtherStocks}$		0.608		0.421		0.836
		(12.42)***		(7.50)***		(32.06)***
R^2	0.02	0.03	0.02	0.03	0.20	0.30
Estimation method	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Instrumental variables	Set 1	Set 2	Set 1	Set 2	Set 1	Set 2

Panel A: Dollar volume
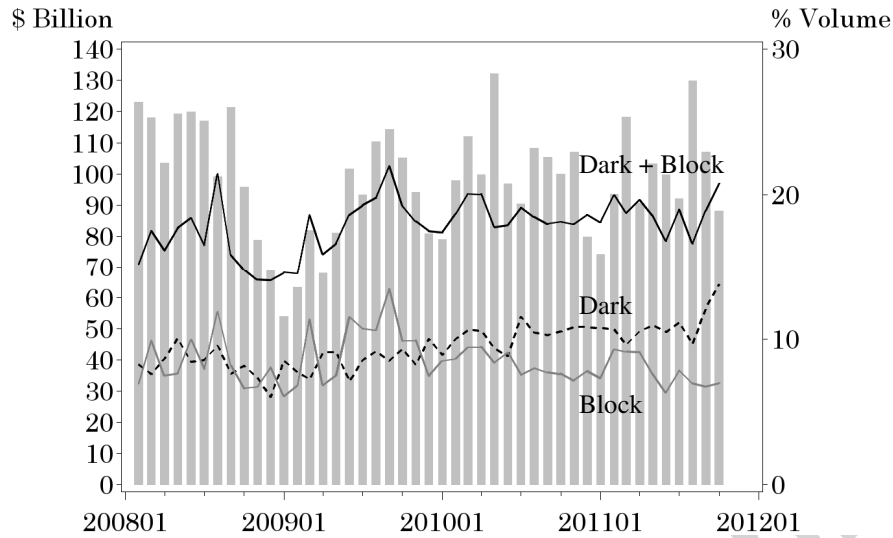
Panel B: Average trade sizes
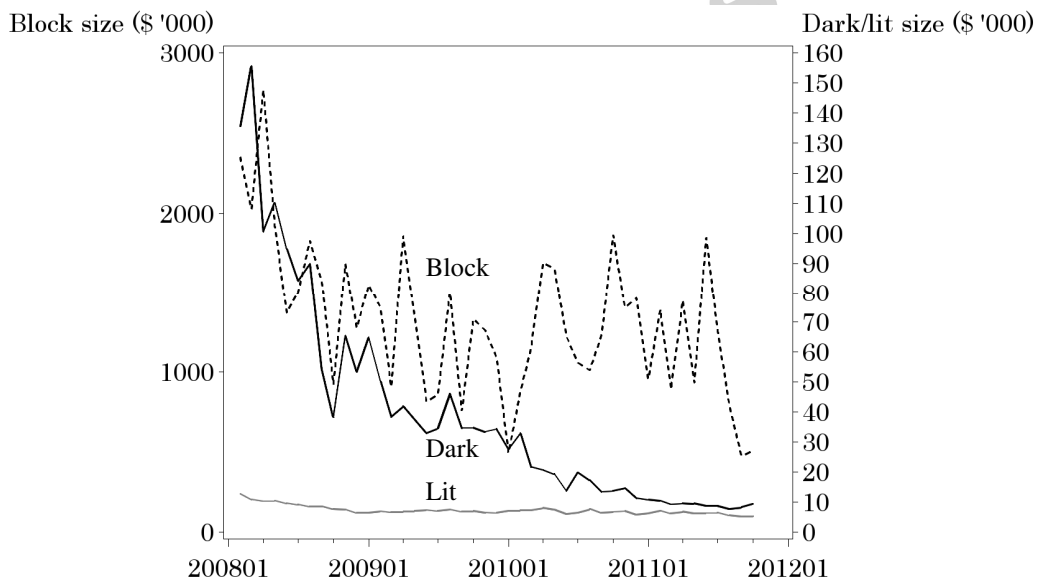
Fig. 1. Dollar volume and average trade sizes. Panel A plots the total dollar volume (in $ billion per month, grey bars) for our sample of stocks (the 500 largest ASX-listed stocks from February 1, 2008 to October 30, 2011). The solid grey and dashed black lines indicate the dollar volume of *Block* and *Dark* trades, respectively, as a percentage of total dollar volume. The solid black line plots the sum of *Block* and *Dark* dollar volume as a percentage of total dollar volume. *Dark* trades are trades executed without pre-trade transparency below block size and *Block* trades are large trades executed without pre-trade transparency. Panel B plots the mean size (in $'000) of *Lit* trades (solid grey line, right-hand side scale), *Dark* trades (solid black line, right-hand side scale), and *Block* trades (dashed black line, left-hand side scale).

Panel A: Dark trading **Panel B: Block trading**
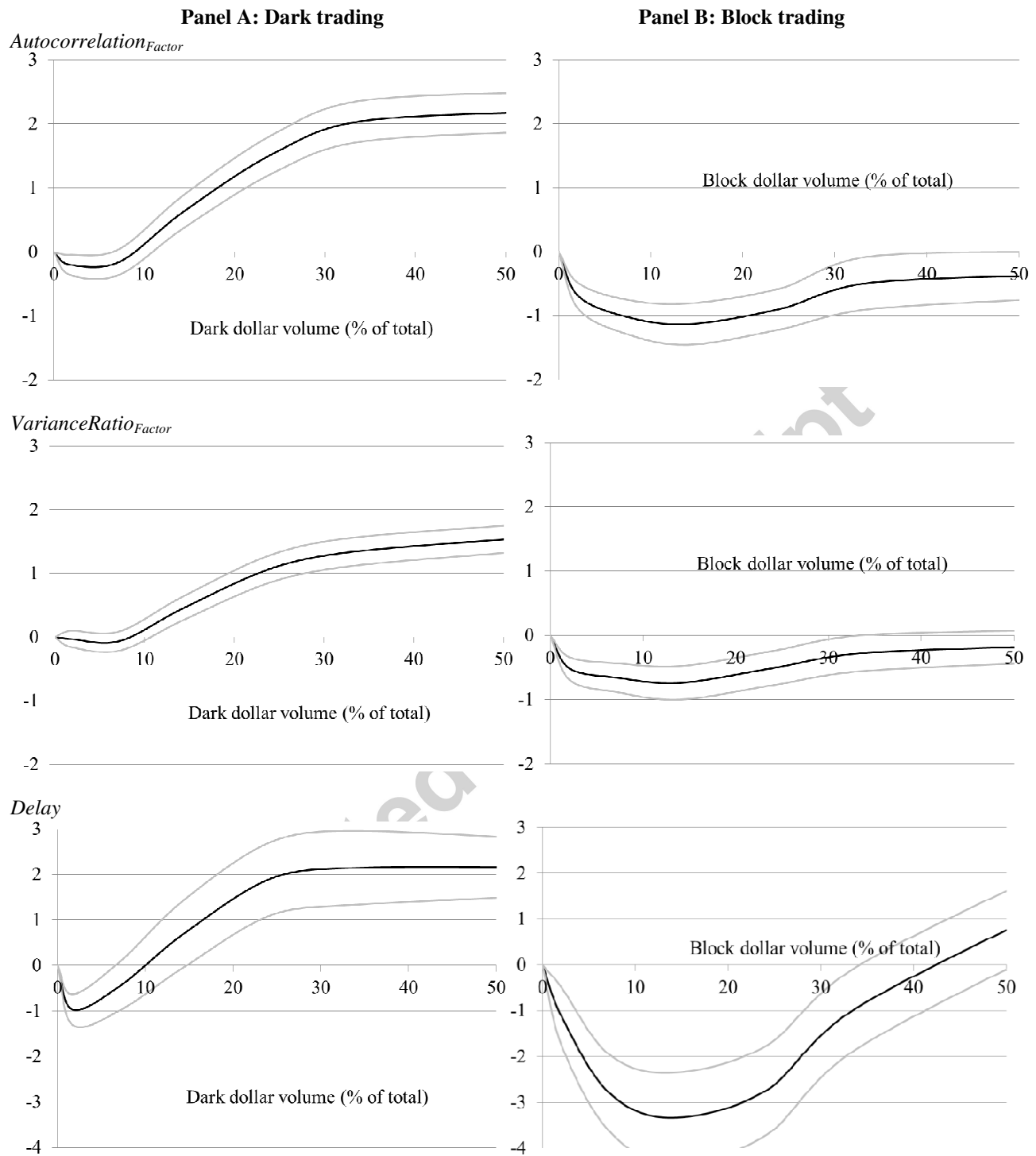
Fig. 2. Effects of dark and block trading on informational efficiency. This figure plots the estimated effects of dark trading (Panel A) and block trading (Panel B) (measured as a percentage of total dollar volume, on the horizontal axis) on three informational inefficiency measures (larger values indicate greater informational *inefficiency*, on the vertical axis). The dark lines plot point estimates and the light lines plot error bounds defined by +/- two standard errors. The estimated effects of dark/block trading are obtained from stock-day panel regressions in which the dependent variables are the informational inefficiency measures and the independent variables are comprised of a set of dummy variables covering various ranges of dark and block trading (0–5%, 5–10%, 10–20%, 20–30%, 30–40%, >40%) and a set of control variables. The sample comprises the 500 largest ASX-listed stocks from February 1, 2008 to October 30, 2011.

54