# Online appendix for “Middlemen in Limit Order Markets”

This online appendix contains two sets of results:

1. Section 1 describes the empirical analysis that serves as input for the model calibration in the main text. Its most relevant part is the diff-in-diff analysis.
2. Section 2 sets up and calibrates the monopoly middleman model. It replicates the analysis in the main text, except for replacing the competitive middlemen with a monopoly middleman.

# 1 Empirical results

This section contains mostly details on the empirical analysis that supports the model calibration in the main text. It is structured into several subsections. The first subsection argues why regulatory change and new market entry created a more friendly environment for high-frequency traders. It also presents the empirical strategy and introduces the sample. The remaining three subsections deal with each of objectives identified in the main text:

1. Identifying a middleman and characterizing its trading behavior.
2. Hard information and middleman participation.
3. Model calibration to identify welfare effect of middleman entry.

## 1.1 Institutional background and sample description

### 1.1.1 Institutional background

The 2007 regulatory change in Europe fostered new market entry and these entrant markets were particularly friendly to HFT activity due to lower fees (to which a *high*-frequency trader is particularly sensitive) and system speed. This section discusses this institutional setting in detail — a setting that is of particular interest to us in our search for a middleman entry event.

The European Union aimed to create a level playing field in investment services when it introduced the *Markets in Financial Instruments Directive* (MIFID) on November 1, 2007. For markets, MIFID created competition between national exchanges and it allowed new markets to enter.

Instinet pre-empted MIFID when it launched the trading platform Chi-X on April 16, 2007, for Dutch and German index stocks.[12] On January 10, 2008, it allowed a consortium of the world

[12] "Chi-X Successfully Begins Full Equity Trading, Clearing and Settlement," Chi-X press release, April 16, 2007.

largest brokers to participate in equity through minority stakes.[13] Before Chi-X, Instinet had successfully introduced the product as "Island" in the U.S. which distinguished itself through fast-execution and subsidization of passive orders (see fee discussion below). Eventually Instinet sold the U.S. license to NASDAQ but kept the international license which led to Chi-X.

Early 2008, our main sample period, Chi-X traded British, Dutch, French, German, and Swiss index stocks. It had captured 4.7% of all trades and was particularly strong in Dutch stocks with a share of 13.6%. Volume-wise, Chi-X overall market share was 3.1% and its Dutch share was 8.4%.

Prior to Chi-X entry, Euronext was by far the main venue for trade in Dutch stocks. Its trading platform ran in much the same way as the Chi-X platform and competition focused on fees and speed (see discussion below). Dutch stocks also traded as ADRs in the U.S. and in the Xetra system run by the German Stock Exchange. They did not yet trade in NASDAQ OMX, Turquoise, or BATS-Europe which entered later on a business model similar to Chi-X: subsidies on passive orders and a fast system.

The broker identified as a middleman in this study was a substantial participant in Chi-X. In our main sample period, it participated in 47.3 million of the 99.2 million Chi-X trades. It was particularly active in Dutch stocks with participation in 4.9 million out of 8.6 million Chi-X trades.

**Fee structure.** Both Chi-X and Euronext operate standard limit order markets. A limit order is a message to buy or sell a number of securities at a pre-specified price. The mechanics of the market are best explained through an example. Suppose all submitted limit orders are for one security. Let us analyze the arrival of a buy order. At any point in time, there is a stack of earlier-submitted sell orders that are waiting at the sell side of the stack of limit orders, the "book". If, on arrival, the buy limit order's price is (weakly) higher than the lowest price on the sell side of the book, it is matched with the lowest-priced sell order (and, if many, with the oldest one) which is removed from the book. The transaction is sealed at the price of the sell order. In this case, the limit buy is

[13]These brokers were: BNP Paribas, Citadel, Citi, Credit Suisse, Fortis, Getco Europe, Ltd, Goldman Sachs, Lehman Brothers, Merrill Lynch, Morgan Stanley, Optiver, Société Générale and UBS (op. cit. footnote 15).

the aggressive side of the trade; the limit sell is the passive side. If, however, the buy order's price is lower it is added to the buy side of the book and therefore effectively becomes a bid price for the security. For a detailed description of these markets we refer to Biais, Hillion, and Spatt (1995).

In our sample period, Chi-X charged substantially lower fees on limit orders than Euronext — the differential is particularly large for passive orders. Chi-X does not charge passive orders on arrival and rebates them 0.2 basis points in case they lead to execution. Chi-X charges aggressive limit orders a 0.3 basis points fee. This type of fee structure is commonly referred to as a maker-taker model. Euronext charges all limit orders that lead to execution, irrespective of their passive or aggressive nature, a fixed fee of €1.20 per trade. For an average size trade of approximately €25,000 (cf. Table 1) this amounts to 0.48 basis points. Highly active brokers benefit from volume discounts that can bring the fixed fee down to €0.99 per trade (0.40 basis points). In addition, Euronext charges an *ad valorem* fee of 0.05 basis points. Limit orders submissions and cancelations are not charged except if, on a daily basis, the cancel-to-trade ratio exceeds 5. In that case, all orders above the threshold get charged a €0.10 fee (0.04 basis points).

As for post-trade costs, Chi-X clears and settles through EMCF which claims to be over 50% cheaper than other European clearing houses including the ones used by Euronext.[14]

**System speed.** In a April 7, 2008 press release Chi-X celebrates its first anniversary. It claims to run one of the fastest platforms in the industry with a system response time (often referred to as "latency") of two milliseconds. This is "up to 10 times faster than the fastest European primary exchange."[15]

Overall, Chi-X appears to be a particularly friendly venue to high-frequency traders. An HFT's key contribution to the market is a technology to reduce the adverse selection cost on the arrival of hard information to a market. To be successful in doing so, a middleman should be able to *instantly* cancel and resubmit a price quote (limit order) and not be charged prohibitively *high fees* on each such action. The Chi-X market seemed to address that need.

[14]See, "EMCF cuts clearing fees", The Trade News, 4/24/08.
[15]"Chi-X Europe Celebrates First Anniversary," Chi-X press release, April 7, 2008.

### 1.1.2 Sample description: data, approach, and summary statistics

**Data.** The main sample consists of trade and quote data on all Dutch nonfinancial index stocks for both Chi-X and Euronext. The sample period is January 10 (just after the last major institutional change to Chi-X became effective, see footnote 15) through April 23, 2008. The financial stocks (FORTIS and AEGON) were removed to mitigate the effects of extraordinary trading in such stocks in the 2007-2008 financial crisis. The quote data are standard and consist of the best bid and ask price and the associated depths at any time point in the sample period (to the second). The trade data contain transaction price, size, and an anonymized broker ID for both sides of the transaction. The broker ID anonymization is done for each market separately and broker IDs can therefore not be matched across markets — say the first market uses 1,2,3,... and the second one uses A,B,C,.... Section 1.2 describes how systematic matching of broker IDs across markets leads to the discovery of a large middleman arriving in the market on July 27, 2007.

In a final analysis we isolate a "treatment" effect of middleman entry in the market through a diff-in-diff analysis (see Section 1.4.1). To this end, the sample is extended with a pre-middleman sample and an untreated benchmark sample. The pre-middleman sample period runs from April 16, 2007 (the day of Chi-X entry) until July 27, 2007 (the day the identified middleman starts its engines). This is a natural sample period as the new-market effect already kicked in (the availability of speed and low fees), yet the identified large middleman was not yet active. There might be other middlemen active in this period but we believe it is unlikely as Chi-X market share was 1-2% and bid-ask spreads did not respond. This changes almost instantly as soon as the new middleman arrives (see, e.g., later discussion on Figure 11). To control for other changes in the trading environment when comparing the pre- and post-middleman sample, Belgian index stocks are analyzed as an "untreated" sample since Chi-X had not been introduced for these stocks. Other than the absence of Chi-X, Belgian stocks traded in the same way as Dutch stocks as Euronext operated the same trading system in its four markets (Belgium, France, Netherlands, and Portugal). In terms of the data, the Belgian sample differs from the Dutch sample only in that it lacks broker IDs on transactions. A list of all sample stocks is included as Appendix A.

Finally, we use an index-futures quote dataset for the Dutch AEX index and the Belgian BEL20 index, both highly active markets. It contains the best bid and ask quotes (to the second).

**Approach.** The analysis was done in two steps (i) to make it feasible (the entire dataset contains roughly 100 million event records) and (ii) to do proper statistical inference. In a first step all variables of interest are calculated for each stock-day. The second step is a panel data analysis on the "box" with all stock-day results. Standard errors are based on time-clustered residuals to account for commonality and heteroskedasticity. The results are presented as weighted-averages across stocks where weights correspond to the local index weights.

**Summary statistics.** Table 1 presents some summary statistics to characterize trading in Chi-X and Euronext. It leads to a couple of observations. First, Chi-X obtained a substantial market share within a year since its inception. Chi-X volume share is 7.3%; its trade share is 11.3%. It was particularly successful for large stocks.

Second, Chi-X contributes substantially to overall liquidity supply. A standard measure of such supply is the quoted bid-ask spread which is defined as the the (log) difference between the lowest ask price in the book and the highest bid price. The Euronext average (time-weighted) half bid-ask spread is 3.77 basis points; the Chi-X half spread is 5.99 basis points. The contribution of Chi-X is best measured by calculation of the inside spread which is defined as the differential between the lowest ask (across the two markets) and the highest bid. The inside half spread is 3.00 basis points, 20% lower than the Euronext spread. The inside spread is, effectively, what investors with "smart routers" pay (Foucault and Menkveld, 2008). These results indicate that Chi-X often features a strictly better price on one side of the market only, i.e., a strictly lower ask or a strictly higher bid. Another important (complementary) measure of liquidity supply is how much can be traded at these best bid and ask quotes, i.e., market depth. This depth is, on average, €98,500 in Euronext and €46,300 in Chi-X. The 20% spread reduction for smart routers therefore only pertains to small orders; for larger orders this is an upper bound.

Third, the cost of adverse selection on a posted bid or ask price is relatively large — it is almost

**Table 1: Summary statistics.** This table provides summary statistics on a sample of Dutch index stocks that trade both in the incumbent market Euronext and in the entrant market Chi-X. The sample runs from January 10 through April 23, 2008. The table reports weighted averages where weights are based on a stock's local index weight. Time-clustered standard errors account for commonality and heteroskedasticity and are reported in parentheses.

| variable (units) | large stocks | small stocks | all stocks |
|---|---|---|---|
| *Panel A: trade activity* | | | |
| Euronext volume (€1000/min) | 412.3 (15.0) | 97.7 (3.4) | 365.6 (13.0) |
| Chi-X volume (€1000/min) | 33.4 (1.4) | 3.5 (0.2) | 28.9 (1.2) |
| Chi-X share volume (%) | 7.5 | 3.4 | 7.3 |
| Euronext #trades (/min) | 16.02 (0.49) | 8.28 (0.22) | 14.87 (0.44) |
| Chi-X #trades (/min) | 2.14 (0.08) | 0.51 (0.03) | 1.90 (0.07) |
| Chi-X share #trades (%) | 11.8 | 5.8 | 11.3 |
| *Panel B: bid-ask spread* | | | |
| Euronext time-weighted quoted half spread (basis points) | 3.44 (0.08) | 5.67 (0.17) | 3.77 (0.09) |
| Chi-X time-weighted quoted half spread (basis points) | 4.07 (0.15) | 17.05 (1.33) | 5.99 (0.29) |
| time-weighted inside half spread[a] (basis points) | 2.67 (0.19) | 4.88 (0.24) | 3.00 (0.19) |
| Euronext time-weighted quoted depth (€1000) | 110.9 (2.0) | 27.4 (0.4) | 98.5 (1.7) |
| Chi-X time-weighted quoted depth (€1000) | 51.3 (1.1) | 17.5 (1.1) | 46.3 (0.9) |
| overall trade-weighted effective half spread (basis points) | 2.83 (0.06) | 4.40 (0.12) | 3.06 (0.07) |
| overall trade-weighted adverse selection, 30 min (basis points) | 2.45 (0.13) | 3.92 (0.20) | 2.67 (0.13) |
| overall trade-weighted realized spread, 30 min (basis points) | 0.38 (0.13) | 0.47 (0.16) | 0.39 (0.12) |
| N=852 (12 stocks, 71 days) | | | |

[a]: defined as the log difference between the lowest ask price (across both markets) and the highest bid price

equal to half the bid-ask spread. A standard empirical measure of adverse selection is the average adverse price movement conditional on a trade. It simply "waits out" the time it takes for the price to reflect the information content of the trade. The measure is defined as:

$$q_{jt}(log(m_{j,t+30min}) - log(m_{jt})), \tag{59}$$

where $q_{jt}$ is a buy-sell indicator (+1 for a buy, −1 for a sell), $m_{jt}$ is the midquote, i.e., the average of the bid and ask price at the time of the $t^{th}$ trade, and $m_{j,t+30min}$ is the midquote 30 minutes later.

The model equivalent is in (32) and (37). The average adverse selection on trades is 2.67 basis points. The *ex post* spread, the distance between the transaction price and the midquote, known as the effective half spread, is 3.06 basis points. The size of adverse selection cost is thus 87% of the effective half spread. The remainder part, the realized spread, is therefore the average gross profit to the passive side of the trade, i.e., the price quote submitter. This spread decomposition is standard in the microstructure literature (e.g., Glosten, 1987).

## 1.2 Identifying a middleman and characterizing its trading behavior

This section searches for a middleman, a machine, and verifies whether it behaves as predicted by the model of Section 2.

### 1.2.1 Caught on tape! a middleman

Pairing broker IDs systematically across markets (1-A, 1-B, ..., 2-A, 2-B, ...) yields one pair that has all the characteristics of a high-frequency trader.[16] Its net position mean-reverts either within the second or, most often, in a matter of minutes (Menkveld, 2013, Table 3). It trades very frequently. It thus matches the SEC definition of a "high-frequency trader."

Table 2 presents trade statistics in support of the conjecture that the identified broker ID pair is a middleman. Panel A reports that on almost half (0.47) of the sample stock-days the middleman's net daily inventory change is zero. On average, this change is €-87,000. Panel B shows that the middleman trades 1.62 times per minute in Euronext and 0.97 times per minute in Chi-X. It trades an average €24,500 per minute in Euronext and €14,800 per minute in Chi-X. The average net inventory change is therefore the same magnitude as the amount it trades in a minute which is evidence of strong inventory mean-reversion. The middleman is a substantial market participant

[16] We started with the most active broker ID in Chi-X, paired all Euronext broker IDs, and immediately discovered the HFT this way. It participated in roughly half of the Chi-X trades and therefore qualifies as the type of middleman this study is focused on.

**Table 2: Caught on tape! a middleman.** This table presents trade statistics for one large middleman. It summarizes its trading in Dutch index stocks from January 10 through April 23, 2008. Time-clustered standard errors account for commonality and heteroskedasticity and are reported in parentheses.

| | large stocks | small stocks | all stocks |
|---|---|---|---|
| *Panel A: middleman inventory* | | | |
| average net change in middleman inventory (€1000) | −104.8 (64.3) | 30.6 (16.4) | −84.7 (55.2) |
| standard deviation net change in middleman inventory (€1000) | 1,088.8 (137.2) | 252.6 (38.0) | 1,009.5 (126.6) |
| fraction of days with zero net change in inventory | 0.38 | 0.57 | 0.47 |
| *Panel B: middleman activity* | | | |
| middleman Euronext volume (€1000/min) | 28.2 (1.6) | 3.3 (0.3) | 24.5 (1.4) |
| middleman Chi-X volume (€1000/min) | 17.2 (1.4) | 1.4 (0.1) | 14.8 (1.2) |
| middleman Euronext #trades (/min) | 1.82 (0.10) | 0.50 (0.04) | 1.62 (0.09) |
| middleman Chi-X #trades (/min) | 1.10 (0.09) | 0.21 (0.02) | 0.97 (0.07) |
| middleman participation rate Euronext trades (%) | 11.2 (0.5) | 6.2 (0.5) | 10.5 (0.5) |
| middleman participation rate Chi-X trades (%) | 51.6 (3.2) | 40.5 (3.1) | 49.9 (3.1) |
| middleman participation rate (%) | 15.4 (0.6) | 8.1 (0.6) | 14.3 (0.5) |
| *Panel C: middleman order types* | | | |
| middleman relative use of passive orders in Euronext (%) | 79.2 (0.6) | 49.5 (2.9) | 74.8 (0.7) |
| middleman relative use of passive orders in Chi-X (%) | 79.4 (0.6) | 83.5 (0.9) | 80.0 (0.6) |
| N=852 (12 stocks, 71 days) | | | |

as it on every other trade in Chi-X (49.9%) and every seventh trade (14.3%) in Euronext. Disaggregation across stock size shows that it participates less in small stocks. It also appears less eager to carry overnight positions for these stocks as the fraction of days with a zero inventory change is 0.59 as opposed to 0.38 for large stocks.

Panel C shows that the middleman trades mostly passively. In 80.8% of its Chi-X transactions it was on the passive side, i.e., it was the price quote that was "consumed" by an incoming limit order (the aggressive side). In Euronext, this number is slightly lower: 74.8%. The standard errors, 0.6% and 0.7% respectively, indicate that this is a structural pattern as the distance to 50% (half passive, half aggressive) is statistically significant.

This predominant use of passive orders is an encouraging result given that our theoretical model generates such passive role endogenously. The discussion surrounding Figures 3 and 4 (bottom-left panel) shows that the middleman is passive in most of its trades. Note that this is not an assumption, but a result as the middleman *does* sometimes hit the seller's ask price, but it just does not happen often.[17] The model calibration, that will be discussed in Section 3.3, generates the same result (see Figure 7). It is intuitive that, in equilibrium, it is the middlemen that stick their neck out with price quotes as they are less likely to be adversely selected.

The middleman's mostly passive trades and its inventory mean-reversion is further consistent with classic market-making models (see, e.g., Ho and Stoll (1981), Amihud and Mendelson (1980), and Hendershott and Menkveld (2014)). In these models, only the middleman issues price quotes *by assumption*. They predict that the market maker skews its quotes opposite to the direction of his net position in order to mean-revert. This could explain why Chi-X quotes, where our middleman is in every other trade, often features strictly better prices on one side of the market only (see the discussion on the "inside spread" in Section 1.1.2).

#### 1.2.2 Middleman ability to process hard information instantly

The theoretical model is built around the unique ability of middleman machines to process hard information instantly; **M** observes $z^h$ and **S** does not. This key assumption is tested by verifying whether middleman quotes for individual stocks reflect one important source of hard information, price updates in the index-futures market. In the standard CAPM model, stock prices consist of essentially two components, the common factor (the market return) and an idiosyncratic factor (a stock-specific return). The idea is that changes in this common factor can be learned from the highly active index-futures market. A middleman machine with its speed advantage is therefore expected to "refresh" its quotes instantly (i.e., in milliseconds) on changes in index-futures prices and avoid adverse selection on this source of information. A natural response would be that an

[17] Admittedly, the model only endogenizes order type at the start of the game when it is the seller against the middleman, not at the end of it when it is, in some cases, the middleman against the buyer (see Figure 2). In the latter case, the middleman is on the passive side by assumption.

exchange provide price-contingent order types for investors to also be able to condition on this information (as suggested in Black (1995)). We will return to this issue at the end of this subsection.

The model's main assumption is tested by comparing the informativeness of both markets' price qoutes and, more specifically, their informativeness on hard information. A direct test of the identified middleman's quotes is not feasible as the data do not have trader ID on quotes (only trades). The information analysis uses Chi-X quotes as a proxy for middleman quotes. This proxy is not unreasonable as (i) Chi-X is the HFT-friendly environment (see fee and speed discussion in Section 1.1.1) and (ii) there is a revealed preference for Chi-X as the identified middleman participated in 49.9% of all Chi-X trades, but only in 14.3% of all Euronext trades.

We believe a cointegration model is the most appropriate econometric approach to measure and compare quote informativeness across markets because informativeness requires one to distinguish the long-term price change (signal) from noise. It is nevertheless useful to start with an intuitive and straightforward analysis on the raw data aimed at verifying whether hard information, index-futures updates, are reflected faster in Chi-X quotes.

**Table 3: Speed of price quote updates: a comparison across markets.** This table compares Euronext and Chi-X in terms of how fast price quotes get updated on changes in the index-futures market. One natural approach is to consider all seconds in the database where the index-futures midquote (the average of the bid and ask quote) changes and count for these events the number of times a markets' stock midquote is adjusted in the same second. To exclude mechanical midquote changes due to executions, the event set is narrowed down to seconds that do not have transactions in the stock one second before, during, or after the index-futures midquote change. The sample consists of trading in Dutch index stocks from January 10 through April 23, 2008. Time-clustered standard errors account for commonality and heteroskedasticity and are reported in parentheses.

| | large stocks | small stocks | all stocks |
|---|---|---|---|
| count of Euronext and index-futures same-second quote changes (/day) | 503<br>(33) | 579<br>(38) | 514<br>(33) |
| count of Chi-X and index-futures same-second quote changes (/day) | 1317<br>(64) | 1162<br>(54) | 1294<br>(60) |
| | | | |
| correlation Euronext quote change and index-futures quote change | 0.31<br>(0.01) | 0.16<br>(0.01) | 0.29<br>(0.01) |
| correlation Chi-X quote change and index-futures quote change | 0.41<br>(0.02) | 0.16<br>(0.01) | 0.37<br>(0.01) |
| | | | |
| 12 stocks, 71 days | | | |

**Simple raw data analysis.** Table 3 shows that Chi-X quotes are more responsive to changes in the Dutch index-futures market than are Euronext quotes. A natural and simple approach is to consider all events where the index-futures midquote changes and count how often a stock midquote adjusts in the same second. This comparison, however, includes "mechanical" stock midquote changes that are the result of executions that knock off stale quotes rather than the result of quote updates. It is for this reason that the analysis further conditions down on index changes that are not accompanied by stock transactions in the same second, one second earlier, or one second later. The results show that same-second Chi-X stock quote updates happen two and a half times more often than Euronext quote updates and their correlation with the index change is more positive. Both these observations are statistically significant.

Panel A in Table 4 provides early evidence that Chi-X quote updates are generally more informative than Euronext quote updates. For both markets, midquote returns are calculated for all inter-trade intervals — the log midquote one second after the previous trade ($t$-1) is subtracted from the log midquote one second before the current trade $t$. These returns are then correlated with the average long-term information revealed through trade $t$ which is proxied for by the transaction price $(t+10)$ minus transaction price $(t-1)$. The correlation thus indicates to what extent a market's quote update "predicts" the information that is about to be revealed in the next trade. This correlation is 0.061 for Chi-X quote updates which is significantly higher than the 0.027 observed for Euronext quote updates. The main drawbacks of this approach are that the wait of 10 periods is arbitrary and, more importantly, the approach does not allow us to verify whether the informativeness differential is due to hard information. This is why we turn to a cointegration approach.

**A cointegration model.** A cointegration model appropriately captures all dynamics in a system of nonstationary price series which features price differentials that are stationary. The cointegatrion model, once estimated, allows for a natural definition of quote informativeness as first proposed in Hasbrouck (1995). The addition of the index-futures allows for a market's quote informativeness to be decomposed into an index-correlated part (hard information) and a remainder part. This decomposition allows us to not only test whether Chi-X quotes are more informative than Euronext quotes, but also whether this differential is particularly strong for hard information (which

**Table 4: Price quote informativeness: a comparison across markets.** This table analyzes to what extent midquote (the average of the bid and ask quote) updates in between trades reveal the information that arrives in the inter-trade intervals. The observation clock therefore runs in transaction time. Panel A correlates log midquote changes strictly in between trades ($t$-1) and $t$ with the information revealed in the trade interval which is proxied by the log trade price for trade ($t$+10) minus log trade price ($t$-1). Panel B is based on a cointegration model that stacks all relevant price series into a single price vector: $p_t = [index_t \quad midquote\_euronext_{t^-} \quad midquote\_chi\_x_{t^-} \quad trade\_price_t]'$, where $t$ runs over the transaction clock, $t^-$ indicates that the quote snapshot is taken one second prior to the transaction, $index$ is the midquote in the local index-futures, $trade\_price_t$ is the transaction price, and $midquote\_X_{t^-}$ is the midquote in market $X$. The cointegration model is: $\Delta p_t = \varphi_1 \Delta p_{t-1} + \varphi_2 \Delta p_{t-2} + \cdots + \beta(A' p_{t-1}) + \varepsilon_t$,

$$\beta' = \begin{pmatrix} 0 & \beta_{22} & \beta_{32} & \beta_{42} \\ 0 & \beta_{21} & \beta_{31} & \beta_{41} \end{pmatrix}, \quad A' = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix},$$

where $\beta A' p_{t-1}$ reflects the presence of two random walks, one associated with the market index and the other with the security's "efficient price" which is naturally defined as: $f_t = \lim_{t\to\infty} E^*[p_{t+k}|p_t, p_{t-1}, \ldots]$, where the asterisk indicates that it is the best *linear* forecast (Hasbrouck, 2007, Ch.8). The extent to which Chi-X and Euronext quotes reveal efficient prices and whether it is the index or the non-index component is established through linear projection of $\Delta f_t$ on the price innovation vector $\varepsilon_t$ where $P_m$ denotes a projection on the first element which captures the market-index innovation (and $P_{-m}$ is its residual), $P_e$ projects onto the second element which is the Euronext quote innovation, and $P_c$ projects onto the third element which is the Chi-X quote innovation. The sample consists of Dutch index stocks from January 10 through April 23, 2008. In parentheses are time-clustered standard errors to account for commonality and heteroskedasticity.

| | large stocks | small stocks | all stocks |
|---|---|---|---|
| *Panel A: correlations based on raw data*[a] | | | |
| corr Euronext midquote ret and long-term price impact of (signed) trade | 0.022 (0.003) | 0.055 (0.004) | 0.027 (0.003) |
| corr Chi-X midquote ret and long-term price impact of (signed) trade | 0.064 (0.004) | 0.048 (0.003) | 0.061 (0.003) |
| *Panel B: cointegration analysis*[a] *(units are basis points squared)* | | | |
| overall efficient price innovation | | | |
| variance eff price innovation in between trades, $\Delta f$ | 8.45 (0.56) | 19.71 (1.47) | 10.12 (0.68) |
| variance eff price innovation correlated with market index, $P_m(\Delta f)$ | 3.51 (0.29) | 6.75 (0.63) | 3.99 (0.33) |
| variance eff price orthogonal to market index, $P_{-m}(\Delta f)$ | 4.95 (0.33) | 12.96 (0.95) | 6.14 (0.41) |
| efficient price innovation correlated with the Euronext midquote return | | | |
| variance eff price innovation in between trades, $P_c(\Delta f)$ | 3.17 (0.20) | 8.33 (0.56) | 3.93 (0.25) |
| variance eff price innovation correlated with market index, $P_m \circ P_c(\Delta f)$ | 0.06 (0.00) | 0.09 (0.01) | 0.06 (0.00) |
| variance eff price orthogonal to market index, $P_{-m} \circ P_c(\Delta f)$ | 3.11 (0.20) | 8.24 (0.56) | 3.87 (0.25) |
| efficient price innovation correlated with the Chi-X midquote return | | | |
| variance eff price innovation in between trades, $P_c(\Delta f)$ | 3.71 (0.21) | 5.35 (0.56) | 3.96 (0.25) |
| variance eff price innovation correlated with market index, $P_m \circ P_c(\Delta f)$ | 0.30 (0.02) | 0.18 (0.03) | 0.28 (0.02) |
| variance eff price orthogonal to market index, $P_{-m} \circ P_c(\Delta f)$ | 3.41 (0.20) | 5.17 (0.55) | 3.67 (0.24) |
| 12 stocks, 71 days | | | |

[a]: based on midquotes strictly inside the inter-trade interval to avoid spurious correlation due to a trade that knocks off the best bid or ask quote

is assumed to the source).

The cointegration model is set up for the following multivariate price vector:

$$p_t = [index_t \quad midquote\_euronext_{t^-} \quad midquote\_chi\_x_{t^-} \quad trade\_price_t]', \tag{60}$$

where $t$ runs over the transaction clock, $t^-$ indicates that the quote snapshot is taken one second prior to the transaction, *index* is the midquote price in the local index-futures, $trade\_price_t$ is the transaction price (in either Chi-X or Euronext), and $midquote\_X_{t^-}$ indicates the midquote price in market $X$. The cointegration model is:

$$\Delta p_t = \varphi_1 \Delta p_{t-1} + \varphi_2 \Delta p_{t-2} + \cdots + \beta(A' p_{t-1}) + \varepsilon_t, \tag{61}$$

$$\beta' = \begin{pmatrix} 0 & \beta_{22} & \beta_{32} & \beta_{42} \\ 0 & \beta_{21} & \beta_{31} & \beta_{41} \end{pmatrix}, \quad A' = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}.$$

The vector error correction term $\beta A' p_{t-1}$ in (61) reflects the presence of two random walks, one associated with the market index and the other with the security's "efficient price". This common efficient price disciplines differentials across Chi-X midquotes, Euronext midquotes, and the trade price series to be stationary with mean zero. Estimation details, the identification of the efficient price innovation, and the linear projections to (i) identify its index and non-index component (used to construct proxies for hard and soft information) and to (ii) find how much of it was already in the Euronext and Chi-X quote updates ahead of the trade are added as Appendix B.

Panel B of Table 4 shows that Chi-X quotes are equally informative overall but significantly more informative on hard information. The proxy for such information is the index-correlated part of efficient price innovations which amounts to approximately 39% (=100%*3.99/10.12). This percentage is in the range of what has been documented for large U.S. firms (see Roll (1988, p.545)). The projections of these efficient price innovations on Euronext and Chi-X quotes reveals that Chi-X quotes are equally informative overall (3.96 vs 3.93 basis points squared). If, however, this differential is decomposed into its index and non-index component, Chi-X is significantly more informative on the index component (0.28 vs. 0.06 basis points squared where the *t*-value on the differential is 9.82). This finding supports the conjecture that middlemen quotes, as proxied by Chi-X quotes, reflect hard information instantly.

The disaggregation across stock size reveals that Chi-X quotes are significantly more informative for large stocks as compared to small stocks. For large stocks, Chi-X quote innovations reveal 17% (3.71 vs. 3.17) more of the inter-trade price innovation than Euronext quote innovations. For small stocks, Chi-X quote innovations reveal 36% less than Euronext quote innovations. The decomposition into index and non-index information reveals that Chi-X quotes are significantly more informative on the index part for both types of stocks, but are significantly less informative on the non-index part for small stocks. For large stocks, the non-index part is not significantly different across markets (although Chi-X informativeness is slightly higher for this part).

The natural and simple solution to the adverse selection friction is to give investors access to hard information contingent order types (as first proposed in Black (1995)). To some extent, these type of orders start to become available to investors. For example, Chi-X allowed its orders to be pegged to price quotes in Euronext (op. cit. footnote 15). To the best of our knowledge, there is no market whose order types condition on, for example, index-futures. From conversations with a NASDAQ official it appears that regulators wish to limit order-type sophistication as small investors might not fully understand complicated order-types. Also, index-futures contingent orders might mitigate some of the advserse selection friction, it will not remove it entirely; correlations are known to be time-varying and, other than index-futures, there are many other sources of hard-information: price quotes in same-industry stocks, foreign exchange, oil and even press releases (there is a hard component to these too that is ground out by text-processing algorithms). Moreover, it might even be efficient to have some agents specialize in designing and maintaining these algorithms especially if, as our theoretical model implies and as the data seem to show, these agents trade mainly passively.

## 1.3 Hard information and middleman participation

This subsection exploits variation in the stock-day panel dataset to study one of the theoretical model's main predictions: investors endogenously choose to involve middlemen more when $\alpha$ is high, i.e.., hard information is relatively important (see Proposition 3 and its Corollary). The natural way to test this prediction is to plot middleman participation in trades against variation in the relative size of hard information. The $\alpha$ proxy for a stock-day is how important index information is for total information on that day for that stock, i.e., the $R^2$ of an intraday single-factor CAPM regression. The regression is implemented for transaction returns where the cointegration approach of Section 1.2.2 avoids an error-in-variables problem associated with straightforward OLS regression. The "intraday $R^2$" is calculated based on the cointegration model estimates as:

$$\frac{var(P_m(\Delta f))}{var(\Delta f)}, \tag{62}$$

where $\Delta f$ tracks the efficient price innovation and $P_m(.)$ is the best linear projection on the market index (see Appendix B).

The scatter plot in Figure 5 provides empirical support for the model's prediction that middlemen participate more when hard information is relatively more important. The intraday $R^2$ shows substantial variation through time as its range is almost 100%. The variation in middleman participation is also substantial as its range is roughly 40%. The scatter cloud indicates a positive correlation which is supported by the regression line. The relationship is economically meaningful as moving from a no-index information day to a full-index information day has the middleman particpate almost 15 percentage points more. This change is somewhat less than the model-implied change (dashed line) when varying only $\alpha$ and keeping the other parameters at their calibrated values (see Section 3.3).

The empirical relationship between middleman participation and $\alpha$ is statistically highly significant as will be revealed through the within correlations that are calculated next.

**Table 5: Correlation daily trade variables.** This table presents correlations for a panel dataset of trade variables that are calculated by stock-day. The variance and quote informativeness measure are based on a cointegration model that identifies the inter-trade efficient price innovation which is labeled $\Delta f$. Model estimates are presented in Table 4. The table reports "between" and "within" correlations which study variable interdependence in the cross-section and through time, respectively. The within correlation is based on time means: $\overline{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$. The between correlation is based on day $t$'s deviation relative to the time mean: $x^*_{it} = x_{it} - \overline{x}_i$. The sample consists of trading in Dutch index stocks from January 10 through April 23, 2008. Standard errors are reported in parentheses (within correlation standard errors account for commonality and heteroskedasticity).

| variable (units) | corr type | middleman participation in trades | variance inter-trade eff price innov | Chi-X share of trades | Chi-X minus Euronext quote informativeness, $\text{var}((P_c - P_e)(\Delta f))$ |
|---|---|---|---|---|---|
| $R^2$ intraday single-factor CAPM[a] (%) | between[b] | 0.48 (0.29) | −0.50 (0.29) | 0.50 (0.29) | 0.59* (0.29) |
| | within[c] | 0.47** (0.06) | 0.13 (0.10) | 0.05 (0.04) | 0.16 (0.08) |
| middleman participation in trades (%) | between[b] | | −0.77** (0.29) | 0.91** (0.29) | 0.76** (0.29) |
| | within[c] | | 0.03 (0.05) | 0.44** (0.05) | 0.17** (0.05) |
| variance inter-trade eff price innov, $\Delta f$ | between[b] | | | −0.74* (0.29) | −0.92** (0.29) |
| | within[c] | | | −0.16** (0.06) | 0.06 (0.06) |
| Chi-X share of trades (%) | between[b] | | | | 0.71* (0.29) |
| | within[c] | | | | 0.01 (0.04) |
| 12 stocks, 71 days | | | | | |

[a]: size of the index component in the efficient price innovation
[b]: based on the time means: $\overline{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$
[c]: based on day $t$'s deviation relative to the time mean: $x^*_{it} = x_{it} - \overline{x}_i$
*/**: significant at a 95/99% level

The panel data correlations reported in Table 5 reveal that, through time, middleman participation correlates with the $\alpha$ proxy, but also with Chi-X market share, and the relative informativeness of Chi-X quotes. The time correlation is picked up by the so-called "within" correlation which removes fixed effects. This correlation is 0.47 for the $\alpha$ proxy and middleman participation with a (robust) $t$-value of 7.83. The correlation with Chi-X share of trades is a significant 0.44. This corroborates the conjectured symbiotic relationship among high-frequency traders and entrant markets (see also Menkveld (2013)). The correlation with how informative Chi-X quotes are relative to Euronext quotes is significantly positive, 0.17, which supports our earlier conjecture (in Section 1.2.2) that Chi-X quotes are more reflective of the middleman's price quotes than Euronext quotes are. It is further worth noting that there is neither a significant correlation between the $\alpha$ proxy and (inter-trade) volatility, nor is there a correlation between volatility and middleman participation. The main result of a large and very significant correlation between the $\alpha$ proxy and middleman participation is therefore unlikely to be spurious in the sense of both series being driven by volatility. The "between" correlations that exploit cross-sectional variation agree with the within correlations except for a significantly negative correlation between middleman participation and inter-trade volatility. In the cross-section, this result is most likely a size effect — small stocks trade less (which implies, ceteris partibus, a higher inter-trade volatility) and middleman participation is less in these stocks (see Table 2).

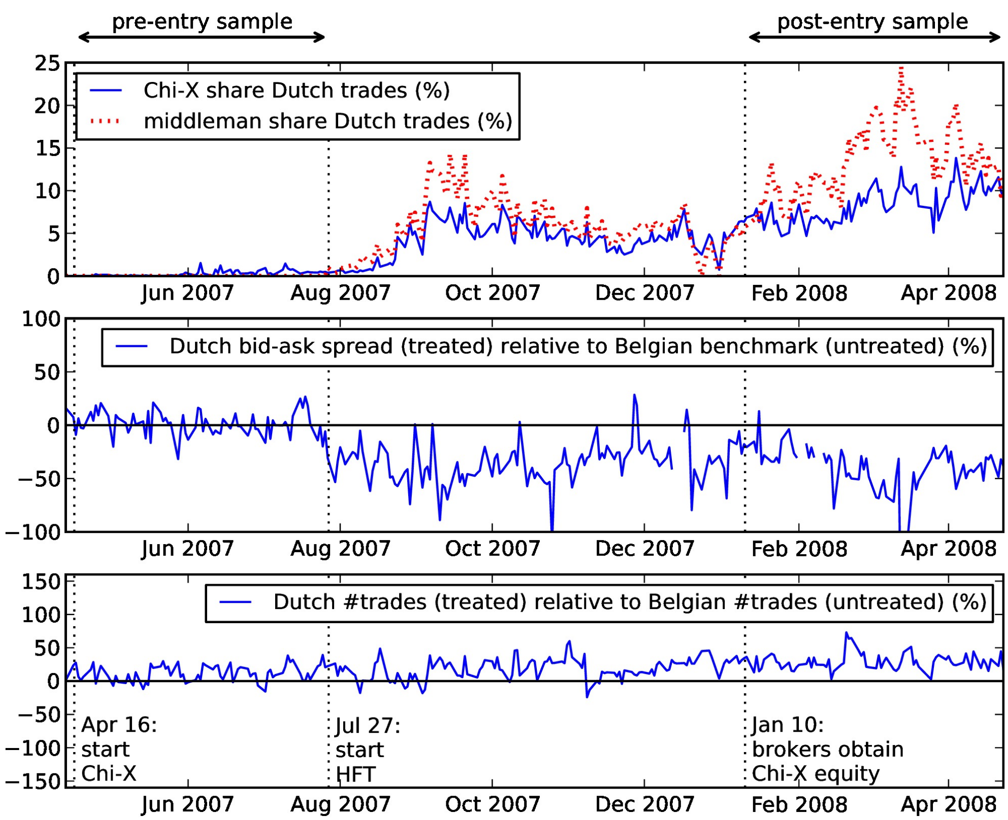
## 1.4 Model calibration to identify welfare effect of middleman entry

### 1.4.1 Diff-in-diff to identify treatment effect

The effect of middleman entry is gauged as the treatment effect that is identified through a diff-in-diff analysis. The first difference is taken across a pre-middleman (April 16 until July 27, 2007) and a post-middleman period (January 10 through April 23, 2008). The second difference is taken across the sample that was treated with middleman introduction, Dutch index stocks, and an untreated sample, Belgian index stocks (which did not experience Chi-X introduction).

Pre-empting the formal diff-in-diff tests, Figure 11 graphs the evolution of the main trading variables around the arrival of the identified high-frequency trader. The top graph shows how the Chi-X share of all trades in Dutch stocks was 1-2% in the early months since its start on April 16, 2007. The real take-off to a double-digit market share happened only after the middleman started its engines on July 27, 2007. The middle graph plots a diff-in-diff (the time diff here is day $t$ minus day 1 of Chi-X introduction) for the bid-ask spread and illustrates the dramatic and immediate decline in the (inside) spread on the day the middleman started; the spread dropped by approximately

**Figure 11: The evolution of Chi-X and HFT trade participation, bid-ask spread, and trade frequency.** This figure plots various trade variables to illustrate how trading responded to the entry of the new trading platform Chi-X on April 16, 2007. The top graph plots the ratio of the number Dutch equity trades in Chi-X and the total number of Dutch equity trades (i.e., Chi-X and Euronext). It also plots the ratio where the denominator is unchanged, but the numerator is the number of trades that the middleman is involved in. The first middleman trade in Dutch equities was on July 27, 2007. The middle graph plots the average bid-ask spread of Dutch stocks relative to the average spread of Belgian stocks. The incumbent market is the same for both sets of stocks (Euronext) but only Dutch stocks are "treated" with Chi-X introduction. The bid-ask spread is the inside spread, i.e., the lowest ask across entrant and incumbent market minus the highest bid across these two markets. The bottom graph plots the same type of graph but now for trade frequency.




30-40% on this day and appears to stay at this reduced level in the remainder of the sample. The bottom graph charts the diff-in-diff for trade frequency and exhibits a relatively modest and slow increase of 10-20%.

The diff-in-diff results of Table 6 formally compare the pre- and post-middleman samples and, most importantly, test for statistical significance of the diff-in-diff effect. The results lead to the following observations. First, volatility changes across both periods are similar when comparing

**Table 6: Diff-in-diff analysis trading variables.** This table documents the results of a difference in difference analysis on various trading variables. The first difference is taken across a pre-middleman (April 16 until July 27, 2007) and a post-middleman period (January 10 through April 23, 2008). The second difference is taken across the sample that was treated with middleman introduction, Dutch index stocks, and an untreated sample, Belgian index stocks (which did not experience Chi-X introduction in this period). The percentage change is calculted based on log differencing. Time-clustered standard errors account for commonality and heteroskedasticity and are reported in parentheses.

| variable (units) | Netherlands/"treated" | | | Belgium/"untreated" | | | diff-in-diff[b] |
|---|---|---|---|---|---|---|---|
| | pre | post | $\Delta$ | pre | post | $\Delta$ | $\Delta\Delta$ |
| 20-min realized volatility (bp/min) | 3.8 (0.1) | 7.1 (0.3) | 60%** (4%) | 4.2 (0.1) | 7.7 (0.3) | 59%** (4%) | 1% (2%) |
| volatility efficient price innovation in between trades, $\Delta f$ (bp) | 2.0 (0.0) | 3.0 (0.1) | 37%** (2%) | 738, 572.0 (732,173.7) | 5.3 (0.2) | 37%** (3%) | 0% (2%) |
| trade-weighted effective half spread (basis points) | 2.54 (0.02) | 3.06 (0.07) | 16%** (2%) | 4.01 (0.06) | 5.33 (0.14) | 32%** (2%) | −15%** (1%) |
| trade-weighted adverse selection, 30 min (basis points) | 1.86 (0.09) | 2.67 (0.13) | 31%** (6%) | 3.19 (0.14) | 5.73 (0.25) | 55%** (6%) | −23%** (7%) |
| time-weighted inside half spread[a] (basis points) | 2.90 (0.02) | 2.17 (0.30) | −17%** (3%) | 5.32 (0.08) | 7.57 (0.20) | 39%** (2%) | −57%** (3%) |
| time-weighted quoted depth (€1000) | 201 (3) | 98 (1) | −64%** (2%) | 31 (0) | 22 (0) | −34%** (1%) | −29%** (2%) |
| #trades (/min) | 10.22 (0.26) | 16.74 (0.51) | 50%** (4%) | 3.92 (0.09) | 5.75 (0.17) | 32%** (3%) | 17%** (2%) |
| #trades after removing middleman's trades (/min) | 10.22 (0.26) | 15.44 (0.46) | 42%** (4%) | 3.92 (0.09) | 5.75 (0.17) | 32%** (3%) | 10%** (2%) |
| volume (€1000/min) | 406 (11) | 392 (14) | 1% (4%) | 67 (2) | 79 (3) | 3% (4%) | −2% (3%) |
| volume after removing middleman's volume (€1000/min) | 406 (11) | 372 (13) | −3% (4%) | 67 (2) | 79 (3) | 3% (4%) | −6%* (3%) |
| #observations 4746, 12+16=28 stocks, 73+71=144 days | | | | | | | |

[a]: defined as the log difference between the lowest ask price (across both markets) and the highest bid price

[b]: defined as Dutch post- minus pre-middleman average minus the equivalent Belgian differential

*/**: significant at a 95/99% level (only applied to differentials)

the treated and untreated stocks. The realized volatility measure, for example, increased by 60% for Dutch stocks and by 59% for Belgian stocks. The differential of 1% is not significant and is economically small, in particular given the magnitude of the overall increase (which is arguably due to onset of the 2007-2008 financial crisis). A cointegration-model-based estimate for the volatility of inter-trade innovations shows that for both the Dutch and the Belgian sample the increased is 37%. To us, this generates a level of comfort with the quality of the match of the main sample and the benchmark sample as we do not expect middleman entry to have an effect on the size of fundamental volatility (which we believe is largely exogenous and driven by the state of a firm and the state of the economy).

Second, on middleman entry, the effective bid-ask spread declined by 15% and the adverse selection cost associated with posted prices declined by 23%. The effective spread itself increased by 32% for the (untreated) Belgian stocks and is arguably due to elevated volatility in the post-middleman period. *Ceteris paribus*, a volatility increase raises the information asymmetry between early and late investors and therefore raises the adverse selection friction. This conjecture is confirmed by the 55% increase in the measured adverse selection cost on posted prices for Belgian stocks. For the (treated) Dutch stocks, these increases are substantially smaller: a 16% spread increase and a 31% increase in adverse selection cost (note that, as observed, the volatility increase for these stocks is just as large as what is observed for the Belgian stocks). The treatment effect (diff-in-diff) is a significant -15% for spread ($t$-value is -8.57) and a significant 23% for adverse selection ($t$-value is -3.04). The treatment effect for *quoted* spread is -57% (see also Figure 11) and for the associated depth on the best quotes it is -29%. A key difference is that the quoted spread measure samples the bid-ask spread at every second whereas the effective spread only samples it at the time of trade. Also, the latter measure is an *ex-post* measure in that it captures the distance to the midquote actually paid by a liquidity demander. This is why we prefer the latter measure.[18]

Third, middleman entry is associated with an increase in trade frequency of 17%. Trade frequency

[18]The reduced depth result implies that lower quoted spread, most likely, overestimates the liquidity supply effect as large trades might require execution at second best bid or ask prices; the effective spread measures the distance from (average) transaction price to midquote and therefore captures an additional price effect if an order executes beyond the best price quote (and "walks" into the book).

increased by 50% for Dutch stocks which is a significant 17% ($t$-value is 7.61) more than the 32% witnessed for Belgian stocks. For completeness, we also calculate the trade frequency change without the "double-counting" due to investor trades that, in the post-entry sample, effectively execute via the middleman (investor→ middleman→ investor) whereas in the pre-entry sample they do not (investor→ investor); all trades that have the identified middleman on one side are counted by a half instead of one. For this trade frequency measure, the treatment effect is a significant 10% increase. Finally, for volume there is no or, for the double-counting corrected sample, a weakly negative treatment effect. Trade size has shrunk with the advent of the middleman.

## 2 Monopoly middleman model

This section sets up and calibrates the monopoly middleman model. The monopoly model is the same as the competitive model, except that instead of many competitive **M** we now have a single **M**. We shall solve for a separating equilibrium in which **M**'s bid is strictly increasing in $z^h$. Then **B** can learn $z^h$ by observing **M**'s bid. This means that the value to **M** of acquiring the asset is the same as it would be in the competitive case. The differences arise (*i*) at the bid stage where **M** will bid less because he enjoys monopsony power, and (*ii*) at the outset when **S** chooses whether to post or to wait for **M**'s bid, expecting the bid to be lower, **S** will be less likely to wait for the bid and as a result, $a$ will be lower than it was in the competitive case.

**Analysis.** We work backwards from **M**'s ask stage, then to **M**'s bidding stage, and then to **S**'s decision of whether to wait or to post.

**M's ask stage.** If **M** has the security, how does he set his ask? With probability $1 - \lambda$, **B** infers $z^h$ from bidding by **M** in period 1. With probability $\lambda$, **B** simply observes $z^h$ himself. This hard information is common knowledge between them in the second period sub-game when **M** owns the security and posts an ask price $p$ for **B** to consider. (Note that $\lambda$ is identified in the data as it affects the no-**M** case.) **B** accepts the offer iff

$$\begin{aligned} y + z^s + z^h &> p \quad \text{with prob. } \lambda, \\ y + z^s + Z\left(p^{\mathrm{bid}}\right) &> p \quad \text{with prob. } 1 - \lambda, \end{aligned} \tag{63}$$

and then **M** gets $p$. If **B** rejects, **B** gets zero and **M** gets $z^s + z^h$. Then the value to **M** of owning the security at $z^{\mathrm{h}}$ and at $Z\left(p^{\mathrm{bid}}\right) = Z$ is

$$V\left(z^{\mathrm{h}}, Z\right) = \max_p \int_{-\infty}^{\infty} \left\{ \begin{array}{c} \lambda\left(p\left[1 - G^s\left(p - y - z^{\mathrm{h}}\right)\right] + \int_{-\infty}^{p-y-z^{\mathrm{h}}} \left(z^{\mathrm{h}} + z\right) g^s(z)\, dz\right) + \\ (1-\lambda)\left(p\left[1 - G^s(p - y - Z)\right] + \int_{-\infty}^{p-y-Z} \left(z^{\mathrm{h}} + z\right) g^s(z)\, dz\right) \end{array} \right\} dF(y). \tag{64}$$

**M**'s incentive to signal to **B** is proportional to $1 - \lambda$. The envelope theorem implies

$$\begin{aligned} \left.\frac{\partial V}{\partial Z}\left(z^{\mathrm{h}}, Z\right)\right|_{Z=z^{\mathrm{h}}} &= \int_{-\infty}^{\infty} (1-\lambda)\left(p^{\mathrm{ask}} - \left(z^{\mathrm{h}} + p^{\mathrm{ask}} - y - z^{\mathrm{h}}\right)\right) g^s\left(p - y - z^{\mathrm{h}}\right) dF(y) \\ &= \int_{-\infty}^{\infty} (1-\lambda)\, y g^s\left(p - y - z^{\mathrm{h}}\right) dF(y) \end{aligned}$$

which is (49) of the main text.

**M's bidding stage.** **S** posts for $x > a$ and waits for a bid for $x \leq a$. **M** will therefore use the conditional distribution for $x$ with CDF

$$\hat{F}(x \mid x \leq a) = \frac{F(x)}{F(a)} \quad \text{for } x \leq a, \text{ with density } \hat{f}(x) = \frac{f(x)}{F(a)}.$$

We consider a separating equilibrium: **S** accepts bid iff $x < p - Z$, in which case **M** gets $V\left(z^{\mathrm{h}}, Z\right)$. Otherwise **M** gets zero. Therefore, since **M** takes $a$ as given, $1/F(a)$ is a multiplicative constant that does not influence the solution. Thus,

$$p^{\mathrm{b}}\left(z^{\mathrm{h}}\right) = \arg\max_{p \leq a + z^h} \left\{\left[V\left(z^{\mathrm{h}}, Z(p)\right) - p\right] \hat{F}\left(p - Z(p)\right)\right\}. \tag{65}$$

**M** controls the beliefs of both **B** and **S**. The FOC (evaluated at $z^{\mathrm{h}} = Z(p) = z$) is

$$0 = F(p-z)\left[\frac{\partial V}{\partial Z}(z,z)\frac{dz}{dp} - 1\right] + [V(z,z) - p]\, f(p-z)\left(1 - \frac{dz}{dp}\right), \tag{66}$$

displayed in (48) of the main text, where

$$\frac{\partial V}{\partial Z}(z,z) = \int_{-\infty}^{\infty} (1-\lambda)\, y g^{s}(p-y-z)\, dF(y),$$

which is displayed in (49) of the main text. The ODE then becomes

$$\frac{dz}{dp} = \frac{[V(z,z)-p]\, f(p-z) - F(p-z)}{[V(z,z)-p]\, f(p-z) - F(p-z)\frac{\partial V}{\partial Z}(z,z)}.$$

**To post or not to post decision by S.** We need to compare the two options that **S** has. In the separating equilibrium **B** knows $z^{\mathrm{h}}$ anyway, either directly or via **M**'s bid.

1. If **S** posts $p^{\mathbf{S}}(x)$, it will be accepted if either an **M** or **B** values the security more, i.e., iff

$$p^{\mathbf{S}} < \max\left(V\left(z^{\mathrm{h}}, z^{\mathrm{h}}\right), y + z^{s} + z^{h}\right). \tag{67}$$

   If he does not sell, he gets

$$x + E\left(z^{\mathrm{h}} + z^{\mathrm{s}} \mid \max\left(V\left(z^{\mathrm{h}}, z^{\mathrm{h}}\right), y + z^{s} + z^{h}\right) < p^{\mathbf{S}}\right).$$

   So, if he posts he solves

$$\begin{aligned} V(x) &= \max_{p} \int_{\left\{p > \max\left(V\left(z^{\mathrm{h}}, z^{\mathrm{h}}\right), y+z^{s}+z^{h}\right)\right\}} \left(x + z^{s} + z^{h}\right) dG^{\mathrm{h}}\left(z^{\mathrm{h}}\right) dG^{s}(z^{s})\, dF(y) \\ &\quad + p \int_{\left\{p \le \max\left(V\left(z^{\mathrm{h}}, z^{\mathrm{h}}\right), y+z^{s}+z^{h}\right)\right\}} dG^{\mathrm{h}}\left(z^{\mathrm{h}}\right) dG^{s}(z^{s})\, dF(y). \end{aligned}$$

2. If **S** does not post, he gets to infer $z^{\mathrm{h}}$ from **M**'s bid, and then chooses to accept or not. In this

case only **M**'s bid is relevant (**B** arrives too late to make a bid). **S** is not adversely selected because **M**'s bid is part of the separating equilibrium. Therefore **S**'s utility in case he decides not to post is:

$$U(x) = \max\left(p\left(z^{\mathrm{h}}\right), x + z^{\mathrm{h}}\right).$$

Then $a$ solves for $x$ the equation

$$U(x) = V(x).$$

# A List of all stocks

This table lists all stocks that are analyzed in this manuscript. It reports the official isin code, the company's name, and the index weight which was used throughout to calculate (weighted) averages. The financial stocks were removed from the sample due to their extraordinary trading patterns in the 2007-2008 financial crisis.

| Dutch local index stocks / "treated" sample | | | Belgium local index stocks / "untreated" sample | | |
|---|---|---|---|---|---|
| isin code | security name | index weight[a] | isin code | security name | index weight[a] |
| NL0000009470 | royal dutch petrol | 28.6% | BE0003565737 | kbc | 23.0% |
| NL0000009538 | kon philips electr | 17.2% | BE0003793107 | interbrew | 13.8% |
| NL0000009355 | unilever | 16.1% | BE0003470755 | solvay | 9.4% |
| NL0000009082 | koninklijke kpn | 10.5% | BE0003797140 | gpe bruxel.lambert | 8.3% |
| NL0000009066 | tnt | 6.8% | BE0003562700 | delhaize group | 7.8% |
| NL0000009132 | akzo nobel | 6.0% | BE0003810273 | belgacom | 6.4% |
| NL0000009165 | heineken | 4.3% | BE0003739530 | ucb | 6.0% |
| NL0000009827 | dsm | 3.4% | BE0003845626 | cnp | 4.2% |
| NL0000395903 | wolters kluwer | 3.1% | BE0003775898 | colruyt | 3.3% |
| NL0000360618 | sbm offshore | 1.7% | BE0003593044 | cofinimmo | 3.2% |
| NL0000379121 | randstad | 1.6% | BE0003764785 | ackermans and van haaren | 3.2% |
| NL0000387058 | tomtom | 0.9% | BE0003678894 | befimmo-sicafi | 3.1% |
| | | | BE0003826436 | telenet | 3.1% |
| | | | BE0003735496 | mobistar | 2.2% |
| | | | BE0003780948 | bekaert | 1.6% |
| | | | BE0003785020 | omega pharma | 1.5% |

[a]: The index weights are based on the true index weights of December 31, 2007.

# B Details on the cointegration approach

This appendix describes in detail the cointegration analysis of Section 1.2.2. The cointegration model is set up for the following multivariate price vector:

$$p_t = [index_t \quad midquote\_euronext_{t^-} \quad midquote\_chi\_x_{t^-} \quad trade\_price_t]', \tag{68}$$

where $t$ runs over the transaction clock, $t^-$ indicates that the quote snapshot is taken one second prior to the transaction, *index* is the midquote price in the local index-futures, $trade\_price_t$ is the transaction price, and $midquote\_X_{t^-}$ indicates the midquote price in market $X$. The cointegration model is:

$$\Delta p_t = \varphi_1 \Delta p_{t-1} + \varphi_2 \Delta p_{t-2} + \cdots + \beta(A' p_{t-1}) + \varepsilon_t, \tag{69}$$

$$\beta' = \begin{pmatrix} 0 & \beta_{22} & \beta_{32} & \beta_{42} \\ 0 & \beta_{21} & \beta_{31} & \beta_{41} \end{pmatrix}, \quad A' = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}.$$

The vector error correction term $\beta A' p_{t-1}$ in (61) reflects the presence of two random walks, one associated with the market index and the other with the security's "efficient price". This common efficient price disciplines differentials across Chi-X midquotes, Euronext midquotes, and the trade price series to be stationary with mean zero. Price changes are assumed to be covariance stationary which implies that they can be expressed as a vector moving average (VMA):

$$\Delta p_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots = \theta(L)\varepsilon_t, \tag{70}$$

where $L$ is the lag operator. The two random walks show up in the coefficient polynomial $\theta(L)$ evaluated at one which reflects the long-term response of prices to an error term impulse. This matrix has rank 2, i.e., the second, third, and fourth row are equal as all three are security prices that in the long-term agree on what the current shock's impact is on the efficient price. A useful econometric proxy for this efficient price is the best long-term linear forecast of prices conditional on all historical price information up until and including time $t$:

$$f_t = \lim_{t\to\infty} E^*[p_{t+k} | p_t, p_{t-1}, \dots], \tag{71}$$

where the asterisk indicates that it is the best *linear* forecast. Hasbrouck (2007, Ch.8) shows that the forecast innovation from ($t$-1) to $t$ for the efficient price is:

$$\Delta f_t = [\theta(1)]_2 \varepsilon_t, \tag{72}$$

where $[.]_i$ indicates the $i^{th}$ row of the matrix in brackets.

The forecast innovations $\Delta f_t$ are a natural measure for the theoretical model's common value changes in between trades and linear projection allows for further analysis of these changes. For ease of exposition, let $P_x(y)$ be the best linear projection of the random variable $y$ on random variable $x$. In regression terms,

$$P_x(y) = x'\beta, \tag{73}$$

where $\beta$ is the coefficient of a standard linear regression of $y$ on $x$. The projections allow for the following analysis.

1. Quote informativeness in between trades is naturally measured by the variance of efficient price changes projected on the quote innovation, e.g., $var(P_e(\Delta f_t))$ where $P_e$ projects on the Euronext quote innovation $[\varepsilon]_2$. In other words, how much information in the inter-trade interval can be learned from obtaining a market's quote update?

2. The variance of efficient price changes (denoted by $\Delta f_t$) and its projection on quote innovations can be decomposed into an index-correlated component (hard information) and an orthogonal component. For example, $var(P_m \circ P_e(\Delta f_t))$ where $m$ corresponds to $[\varepsilon]_1$ indicates how much of Euronext quote informativeness reflects the index innovation. If Euronext quote updates are uncorrelated with index innovations this measure is zero. $P_{-m}$ is defined to be the orthogonal part, i.e., $y = (P_m + P_{-m})(y)$ by construction.