Market Microstructure PhD Course:
Course Notes for Empirical Work

Professor Talis J. Putnins

CONTENTS

1. EMPIRICAL SETTING STUDIED IN THIS COURSE

To illustrate and apply various microstructure models and metrics we have chosen a particular empirical setting – a sample of stocks traded on the Nasdaq OMX Nordic Stock Exchanges (Stockholm, Helsinki, Copenhagen) during a two month period in 2008. Two interesting features of this setting are: (i) several stocks are cross-listed on different Nordic exchanges, which allows for some interesting analysis; and (ii) a subset of stocks switches from post-trade transparency of broker IDs to post-trade anonymity of broker IDs during the sample period, creating a natural microstructure experiment. In addition to these interesting features, the chosen exchanges are automated electronic limit order book markets, and therefore similar in structure to many of the world's major financial markets. This section describes some of the institutional detail required to apply the microstructure techniques covered in this course.

Institutional details of the Nasdaq OMX Nordic Stock Exchanges (Stockholm, Helsinki, Copenhagen)[1]

- Limit order market with price-internal(member)-display-time or price-display-time priority depending on the order types.
- Trading hours:
  - Stockholm: 09:00 – 17:30 local time (CET)
  - Copenhagen: 09:00 – 17:00 local time (CET)
  - Helsinki: 10:00-18:30 local time (EET = CET +1)
  - CET time is 2 hours ahead of GMT during daylight savings (strictly speaking this is Central Europe Summer Time, CEST, which is approximately April to October), and 1 hour ahead otherwise.
  - Note: during the last 10 minutes orders are not matched (pre-close period) and a batch auction takes place at a random time in the 30-second period between 4.5 and 5 minutes into the pre-close period. Therefore, allowing for open/close, all markets should be in continuous trading between 09:10 and 16:50 (CET).

[1] Technically, Nasdaq OMX Iceland is also part of the Nasdaq OMX Nordic Exchanges group, but it is considerably smaller than the other three exchanges and tends to have a different set of exchange members to the other three exchanges (which have most members in common).

- Open and close with a call auction
- Main stock indices:
  - OMX Stockholm 30 is a value-weighted index of the top 30 stocks listed in Stockholm by trading activity
  - OMX Helsinki 25 is a value-weighted index of the top 25 stocks listed in Helsinki by trading activity
  - OMX Copenhagen 20 is a value-weighted index of the top 20 stocks listed in Copenhagen by trading activity
- Trading currencies: EUR (Helsinki); SEK (Stockholm); DKK (Copenhagen)
- Cross-listings between the Stockholm-Helsinki-Copenhagen exchanges during the period May 2008 – May 2009: see Table 1.

Table 1
Cross-listed stocks between the Stockholm-Helsinki-Copenhagen exchanges during the period May 2008 – May 2009

Company	Primary listing	Other listings
Nordea	STO (NDA.ST)	HEX (NDA1V.HE)
TeliaSonera	STO (TLSN.ST)	HEX (TLS1V.HE)
Nordea	STO (NDA.ST)	CPH (NDA.CO)
SAS	STO (SAS.ST)	CPH (SAS.CO)
Stora Enso R	HEX (STERV.HE)	STO (STEr.ST)
TietoEnator	HEX (TIE1V.HE)	STO (TIEN.ST)
Nokia	HEX (NOK1V.HE)	STO (NOKI.ST)

- Tick size:
  - Helsinki: EUR 0.01 for all stocks regardless of price
  - Stockholm: Tick size tables (depends on price band, and different tables for different stocks)
  - Copenhagen: Tick size tables (depends on price band, and different tables for different stocks)
- Transparency:
  - Pre-trade: Until 13 March 2006 all three exchanges displayed the broker ID in the order book; since that date broker IDs for orders resting in the limit order book are not publicly displayed. The limit order book, containing prices and volumes, is publicly available.
  - Post-trade (see Table 2):

- Before 2 June 2008: Immediate post-trade transparency on all exchanges (each trade was printed to the tape in real time with price, volume, broker IDs and trade type)
- Between 2 June 2008 and 14 April 2009: Post-trade anonymity for all stocks on Helsinki and top five stocks on Stockholm shown in Table 3 (broker IDs are not disseminated in real time but are available after the close each day). Other stocks in Stockholm and all stocks in Copenhagen retain post-trade transparency.[2]
- After 14 April 2009: Post-trade anonymity retained only for top five stocks on Helsinki, shown in Table 4. All other stocks back to post-trade transparency.[3]

Table 2
Post-trade anonymity

This table reports whether broker IDs are disclosed in real time on trades.

Exchange	Before 2 JUN 2008	2 JUN 2008 – 14 APR 2009	After 14 APR 2009
Helsinki	Transparent	Anonymous	Top 5: Anonymous Others: Transparent
Stockholm	Transparent	Top 5: Anonymous Others: Transparent	Transparent
Copenhagen	Transparent	Transparent	Transparent

Table 3
Five Stockholm stocks that switch to post-trade anonymity on 2 June 2008

Company	In OMXS30 index
Ericsson B (ERICb.ST)	Yes
TeliaSonera (TLSN.ST)	Yes
Volvo B (VOLVb.ST)	Yes
Nordea SEK (NDA.ST)	Yes
H & M B (HMb.ST)	Yes

[2] See https://newsclient.omxgroup.com/cds/DisclosureAttachmentServlet?showInline=true&messageAttachmentId=262624

[3] See https://newsclient.omxgroup.com/cdsPublic/viewDisclosure.action?disclosureId=358365&lang=sv

Table 4
Five Helsinki stocks that keep post-trade anonymity after 14 march 2009

Company	In OMXH25 index
Nokia (NOK1V.HE)	Yes
Fortum (FUM1V.HE)	Yes
UPM-Kymmene (UPM1V.HE)	Yes
Sampo (SAMAS.HE)	Yes
Stora Enso R (STERV.HE)	Yes

- Order types and attributes that can execute during continuous trading:
  - Market
  - Limit
  - Iceberg order (only part of the order is displayed)
  - Pegged order (pegged to the midquote or a certain number of ticks above/below the best ask or bid – if within the spread it is a dark (non-displayed) order type but must execute at a valid (full) tick increment)
  - Minimum quantity / fill-or-kill – only executes if the minimum quantity can be filled
  - Hidden (non-displayed) orders – if an order meets the minimum size requirement
  - Nordic@Mid – a separate continuous crossing system for executing at the midquote (half-tick increments allowed) – no pre-trade transparency, but upon execution trades are reported to the public data feed in real time.
- Smart Order Routing – if the Nasdaq OMX Nordic exchanges do not have the European best quote at the time a marketable order is submitted (i.e., a better quote exists at ChiX, BATS Europe, Turquoise, Burgandy or Oslo Bors) and the member has enabled Smart Order Routing, the exchange will automatically route the order for execution at the venue with the best quote (via an "introducing broker") and then a "mirroring trade" will be executed on the Nordic exchange between the participant that submitted the order and the "introducing broker". This way the client gets to trade at the European best quotes but on the Nordic exchange and the "introducing broker" has a net zero position after executing the trade on the away market and the offsetting trade on the Nordic exchange.

2. WORKING WITH MICROSTRUCTURE DATA

This section provides instructions for sourcing the data, and basic guidance on how to work with large microstructure datasets and prepare them for analysis.

2.1 Obtaining the data

See separate document on “Task 1”.

2.2 Tips for working with large datasets

- Keep only the observations and variables that will be necessary in further analysis.
- Consider which data format will save space. For example, in our analysis we need currency quotes sampled at a 1-second frequency – the data in “Intraday 1 second” format is approximately 500MB whereas in “Time and Sales” format it is only 20MB, and either format can be used to achieve the same analysis. The difference arises because the Time and Sales observations are less frequent than one per second on average.
- Use multiple datasets to cut down on the number of missing or repeated observations. For example, do not store low-frequency variables (e.g., Industry) together with high-frequency data (e.g., trades and quotes) because the repetition of industry will unnecessarily inflate dataset size. Similar logic applies when making data requests.
- Avoid sorting data whenever possible (sorting is cpu/memory/time intensive). For example, to obtain a lead term merge a dataset with itself setting FIRSTOBS=2. For datasets that will be called upon often store them sorted in the most commonly used order.
- Set SAS operating parameters MEMSIZE (in SAS configuration file) and SORTSIZE (using OPTIONS command) to give SAS access to sufficient memory.
- When estimating models with SAS procedures (‘proc …’) keep the datasets down to a size that comfortably fits in RAM. To do so you might consider writing a loop (using a macro) that extracts one stock or one day of data at a time, estimates the model, saves the results and then repeats the process. Even with the overheads associated with the loop the time saving can be HUGE.

- Avoid unnecessarily printing of output/results – many SAS procedures automatically output a lot of additional results, often including resource-intensive graphs etc. Use 'ods _all_ close;' at start to close the default output of additional results to html files. Use 'noprint', 'plots=none', etc to turn off default output from various procedures. You should not turn off output within a procedure when you want to save certain results tables or charts (e.g., with ODS statements).
- Develop and test code on a subset of observations (e.g., use 'options obs=1000000)') and only once you are convinced it is running well, set obs=max and run on the full size datasets. Be careful – depending on the analysis a limited number of observations can cause errors that would not be present if the full datasets were used.
- Install a solid state / flash drive and set SAS's working directory to that drive. Often reading/writing data to a hard disk is the bottleneck in speed, not CPU.

2.3 Common data filters, errors and outliers

- Filter for negative or zero prices and volumes. For example, in some cases a missing bid quote gets replaced with a bid of zero, which can cause incorrect spreads.
- Filter for excessively high prices. Sometimes decimal places are incorrectly missed and $45.305 can turn into $45305.
- Crossed (negative) spreads often indicate that a market is not in continuous trading and therefore midquotes, spreads and volumes are not meaningful.
- The order of microstructure data is important and many errors can arise from data being out of sequence. For example, when a market order executes against one or several resting limit orders and in doing so consumes all of the volume at the best bid or ask the data should show one or several trades and change in the best quotes all identically timestamped. Getting the sequence of events correct is essential to knowing the correct quotes that were prevailing at the time of the trades, which in turn is required to correctly calculate metrics such as effective spreads. Other reasons for out of order data include lack of synchronization of the clocks recording timestamps from different venues (or simply separate trade and quote datasets) that are subsequently consolidated, or transmission delays between the trading venue and database operator. It is

a good idea to record the order of the original data (when importing) so that after various manipulations and sorts you can restore the original sequence (which in many cases is helpful).

- Identifying and dealing with outliers:
  - Output outliers/errors to a separate dataset and inspect them. Do not simply throw them away – understanding why they occurred is important to eliminate other errors/problems that have not yet been identified (and if the problem can be solved the observations may need to be included in the main sample).
  - Quick checks for outliers – sort on the variable of interest and examine the largest and smallest values.
  - Once you are convinced you have valid (not errors) extreme observations you may consider winsorisation of the relevant variables (setting extreme values to the 1st and 99th percentile values, for example).

2.4 Assigning trade initiator

- In a continuous auction market the trade initiator is usually defined as the counterparty to a trade that submitted their order last. For example, when a market order executes against a resting limit order the market order submitter is the trade initiator because their order was submitted after the limit order. If the market order is a buy then the trade is said to be buyer initiated. Knowing the trade initiator is required for several microstructure metrics and models but is rarely provided in datasets and therefore must be inferred (e.g., signed dollar volume = dollar volume of buyer-initiated trades - dollar volume of seller-initiated trades).
- One of the most common approaches to identifying the trade initiator is the Lee and Ready (1991) algorithm and variants of it (still widely used despite being somewhat inaccurate in some settings). The core part of the algorithm is the 'midquote test': if a trade occurs at a price above (below) the prevailing midquote it is classified as buyer (seller) initiated. In some settings this classification rule is more accurate when quotes are lagged by a few seconds. Trades that occur at the midquote can be classified using the pattern of past trade prices (the 'tick test'). Alternatively, trades at the midquote are

sometimes treated as unable to be classified because classification based on past trade prices is often less accurate than the midquote test.

2.5 Other issues for consideration when working with microstructure data

- Currencies
- Time zones
- Opening and closing auctions
- Off-market trades and other sources of trades printed to the tape
- Corrections and nonstandard conditions recorded in qualifiers

2.6 Preparing the data

See separate document on “Task 2”.

3. LIQUIDITY

A fundamental part of many areas of finance is liquidity, but what is liquidity? Many authors have recognized the difficulty in defining liquidity. Maureen O'Hara, in her book on market microstructure, drew an analogy between liquidity and pornography because of the difficulty in defining liquidity. "You know it when you see it". She defines a liquid market to be one where "buyers and sellers can trade into and out of positions quickly and without having large price effects". Perfectly liquid is easy to define as since it refers to an item that can be bought or sold at the same price in whatever quantity. Perfectly illiquid is also easy to describe since it is something that could not be bought or sold at any price. Between the two extremes results in liquidity levels that are relative to each other and depend upon the metric being used to measure liquidity.

There are four conditions that Black (1971) identified as being needed to have a liquid market: (1) bid and ask prices to be always available for small quantities, (2) the difference between bid and ask prices to be small, (3) large amounts of stock could be bought or sold over long periods of time at prices not very different from the current price and (4) immediate sales/purchases of large blocks of stock could be sold/purchased only at a discount/premium that reflects the size of the block.

Inherent in the above and any liquidity definition are many dimensions. Harris (2002) identified four interrelated liquidity dimensions: depth, width, immediacy and resiliency. Depth is defined as the volume that can be traded without having a large effect on price. Width is the difference between the fundamental price and the transaction price. Immediacy is the speed of trade execution and resiliency is the length of time for the price to move back to equilibrium after a large trade. These are not necessarily the only ways of parsing liquidity into dimensions.

The seminal literature on liquidity defined the three main dimensions of liquidity to be spread, depth and resiliency. Kyle (1985) modeled continuous trading by a series of sequential auctions with three different parties. The market maker could observe the total traded and prices but did not know whether the trades were from informed traders or from noise traders. To compare the 'liquidity properties' in the continuous auction equilibrium market that Kyle was describing, he used the term tightness (instead of spread) to refer to the cost of turning around a position over a short period of time. Depth refers to the ability of the market to absorb quantities without having a large effect on price. Resiliency refers to the speed with which prices tend to converge towards the underlying liquidation value of the commodity. Kyle also referred to resiliency as the measure of the rate at which prices bounce back from an uninformative shock.

For the liquidity section of this course, the market microstructure techniques that we will examine are divided into three broad categories that are similar to the different dimensions but are strictly classified into metrics that can be easily estimated. The next section (Section 3.1) examines spreads and goes through quoted, effective and realized spreads. The section after that (Section 3.2) investigates depth while the last section (Section 3.3) examines price impact. Within each section are some related measures as well as some of the background about what limitations may be present when estimating and using these measures. We also include references to some metrics that try to measure liquidity across multiple dimensions.

3.1 Spreads

3.1.1 Quoted spreads

The relative quoted bid-ask spread at time *t* (stock subscript suppressed) is:

$$QuotedSpread_t = \left(Ask_t - Bid_t\right)/m_t \tag{L.1}$$

where Ask_t and Bid_t are the best ask and bid quotes at time *t*, and $m_t = \left(Ask_t + Bid_t\right)/2$ is the midquote at time *t*. The quoted spread indicates the round-trip transaction costs of small market orders that trade against the limit orders resting in the limit order book. A common way to measure the quoted spread at the stock-day level is by calculating the time-weighted average of the quoted spread throughout the day.

Quoted spreads can also be examined continuously during the trading day or within different windows during the day. The first market microstructure paper to show an intraday pattern in spreads was McInish and Wood (1992). They measured a minute-by-minute spread and showed a “crude reverse J-shaped pattern” where spreads were largest at the beginning and end of each day. They identified an inverse relation between spreads and measures of activity such as the number of trades and the number of shares per trade. A different pattern was observed by Chan, Christie and Schultz (1995) whereby the spreads narrowed near the close of trading. The differing market structure (specialists on the NYSE versus market makers on the

NASDAQ) was suggested as an explanation for the difference between market structures.

There are several reasons to expect a positive bid-ask spread. One is the presence of informed traders and the adverse selection costs they create. For example, in the Glosten and Milgrom (1985) model, traders with superior information lead to a positive bid-ask spread even when the specialist is risk-neutral and makes zero expected profits. The bid-ask spread compensates the liquidity provider for losses incurred when he/she trades against the informed traders - adverse selection costs. Posting bid and ask quotes is like providing two free options to market participants that can choose to trade at those quotes. Informed traders exploit these free options when they are in the money – when the bid is above the fundamental value or when the ask is below the fundamental value.

Another reason a liquidity provider would set a positive bid-ask spread is inventory risk. When a liquidity provider accommodates the liquidity demands of other traders (either by selling them stock or buying stock from them) the liquidity provider's inventory position becomes non-zero (either positive or negative depending on what direction trade was accommodated). Holding positive or negative inventory is risky because the fundamental value can change before the inventory position is reversed, causing a loss or a gain on the position. If liquidity providers are risk adverse they would charge a price for bearing inventory risk – this is known as the inventory costs component of the spread.

An entire 'industry' of bid-ask spread component models has developed to try to estimate the theoretical components of the spread and they are loosely defined as either serial covariance related or trade indicator related. These models are not the focus of this course but we introduce awareness of the components as a way to be mindful of different directions that the literature took over time. The theoretical components have been defined as adverse selection costs, inventory management costs and order processing costs. These components are dependent on an institutional setup whereby there is a specialist or market maker that incurs inventory costs for which she is compensated. Failing that, such as when there are multiple market makers without an obligation to hold inventory, there is no rationale for having an inventory cost component. The basis of the order processing cost component is also interesting since over time this cost has become less and less important with automation and improvements in technology. In fact, the decline of the specialist model has left this literature a bit weak since the theoretical foundations of the decomposition models rely on assumptions that are not valid in almost all of today's stock exchanges.

The adverse selection component, a measure of the compensation for trading with someone with information is really the only component that can vary during the trading day so it is not surprising that variations in the spread itself could be assumed to be being caused by variations in adverse selection. We make that loose assumption if needed while we focus on the empirical evidence.

3.1.2 Effective spreads

The quoted spread can over- or under-state execution costs for liquidity demanding trades when trades can execute within or beyond the best prevailing quotes. Trades can execute within the quoted spread for a number of reasons, including: (i) execution against hidden orders; (ii) dealers/market makers giving orders price improvement relative to the prevailing limit order book quotes (or brokers that internalize client orders); and (iii) trades negotiated between two counterparties off market. Trades can execute at prices beyond the best prevailing quotes for a number of reasons, including: (i) large market orders can execute all of the available volume at the best price step and then continue 'walking' through the limit order book and executing at limit orders at the next price step, before the market order is filled; and (ii) trades negotiated away from the market.

The effective spread accounts for the fact that liquidity demanding trades do not always execute at the best quotes by comparing the trade execution price to the prevailing midquote (implicitly used as a proxy for the fundamental value). The relative effective spreads for a trade that occurs at time t (stock subscript suppressed) is:

$$EffectiveSpread_t = 2q_t\left(P_t - m_t\right)/m_t \tag{L.2}$$

where q_t is a trade direction indicator (q=+1 for buyer-initiated trades and q=-1 for seller-initiated trades), P_t is price at which the trade executes, and m_t is the midquote at time t. The factor of 2 makes the effective spread interpretable as the execution cost of a roundtrip liquidity demanding trade, and makes the scale comparable to that of quoted spreads. Effective spreads tend to be smaller than quoted spreads because trades more often occur within the spread than beyond the best quotes. A common way to measure the effective spread at the stock-day level is by calculating the dollar volume-weighted average of the effective spreads for each trade within the stock-day.

3.1.3 Effective spreads

Trades typically have positive price impact, i.e., quotes tend to move up following buyer-initiated trades and vice versa for seller-initiated trades. Comparing a trade's execution price to the midquote prevailing at the time of the trade does not take the trade's price impact into account and therefore (arguably) overstates the trade's true 'execution cost' when execution cost is defined as the price at which a transaction is conducted, relative to the true value of the asset. Similarly, the spread is often interpreted as the revenue earned by a liquidity provider for facilitating a trade, but quotes typically move against the liquidity provider after facilitating a trade then the effective spread will overstate the revenue earned by the liquidity provider. To account for these effects we can use an alternative spread estimate, the relative realized spread, which takes into consideration the price impact of trades:

$$RealizedSpread_t = 2q_t\left(P_t - m_{t+5\min}\right)/m_t \tag{L.3}$$

where q_t is a trade direction indicator (q=+1 for buyer-initiated trades and q=-1 for seller-initiated trades), and m_t is the midquote at time t. In contrast to the effective spread, the realized spread compares the trade price to the midquote at some future point in time once the trade's price impact has been realized (typically five minutes in the future). Because trades tend to have positive price impact, realized spreads are usually smaller than effective spreads on average. A common way to measure the realized spread at the stock-day level is by calculating the dollar volume-weighted average of the realized spreads for each trade within the stock-day.

The use of the trade direction indicator results in dependency on accurate estimation of this variable. In order driven markets relying on an estimate of this measure should not be too much of a problem if the data are correct. Unlike a quote-driven market, there should be little uncertainty about whether the buyer or seller triggered the trade. Simple checks of where the trades occur within and around the bid-ask spread will identify whether the trade indicator is accurate or whether more investigation into the data is needed.

3.2 Depth and the slope of the limit order book

Depth represents the extent of the ability to trade at the prevailing prices. Quoted depth represents the amount of depth that is visible in the quotes at the best ask and bid prices. The actual available depth may be larger if there are hidden orders (only available in some markets). Depth is one of the activity variables and it represents activity that can occur without a price change. Broader activity measures include volume or number of transactions and these activity measures portray the level of actual activity which may be larger than only the quoted depth.

An interesting natural experiment has occurred in various financial markets whereby reductions in tick sizes (such as on the NYSE from eighths to sixteenths and then decimalized) resulted in a much lower level of depth at the prevailing quotes. In fact, Goldstein and Kavajecz (2000) found that the depth in the entire limit order book declined after the change from eighths to sixteenths. Bourghelle and Declerck (2004) examined the changes in the tick size on Euronext (some stocks moved to larger tick sizes while others moved to smaller tick sizes) and found that smaller tick sizes resulted in smaller levels of depth, thereby reducing the level of order exposure and lowering transparency not just at the best quotes but even into the limit order book.

So is depth even relevant on its own? It provides the lower bound of a measure of available liquidity. The reduction in tick sizes alludes to the interrelationship between spreads and depths that is not captured when only looking at one variable. Depth is inextricably linked to spreads, since a small depth level results in quoted spreads that are relevant to only a small amount of trading.

Expanded totals of depth that include several steps of depth within the limit order book can provide a more comprehensive picture of the available volume of trading that reflects the additional cost of trading as well. These expanded measures are actually measuring the tradeoff between price and available volume. Engle and Lange (2001) created a measure, VNET, which directly measures the depth of the market corresponding to a particular price deterioration. They construct the measure using the excess volume of buys or sells associated with a price movement. Their measure reflects the additional volume that is traded above the quoted depth and to some degree this can be loosely considered the time dimension, or resiliency or patience, involved when people are willing to wait for some trading rather than simply take all the orders available in the limit order book.

Another way of measuring depth can include the extent of the trading that would be needed to transact a certain dollar amount relative to the midpoint of the bid and ask price. For example, if a trade of $50,000 were desired, then what level of shares would be available as you 'walked' up the book? The total number of shares that instantly completed the $50,000 buy or sell reflects the liquidity available in the market. If all $50,000 were available at the best ask, this would be a lower price than if additional shares in the limit order book were needed. To make the number of shares meaningful, it would have to be relative to a base which could be the $50,000 bought or sold at the bid-ask spread midpoint. Similarly, to sell $50,000 worth of shares at the best bid would indicate a certain level of bid depth. However, if all $50,000 were not available at the best bid, then the total number of shares sold to receive $50,000 would be larger. In either case, the premium or discount from the total value of a theoretical 'midpoint' price purchase or sale could be a relevant metric.

Standardizing across stocks is also desirable but price levels may not be the most appropriate and a better indication may be a percentage of free float since this is often a concern for fund managers when they are looking at investments. In fact, the Standard and Poors indices recently changed to a free float weighting from a market capitalization weighting to reflect available liquidity in the market. In order to adjust the metric to be comparable across stocks a percentage of the free float could be selected, say 0.01%. Then, for each company we could calculate the number of shares that constituted 0.01% and then use that number similar to the $50,000 in the prior paragraph.

The limit order book provides the minimum available trading that is visible at a point in time. It reflects the trade-off between the price discount or premium that is needed if you want additional volume above that shown at the best quotes. As you step into the book, the increase in the available volume can be measured as the slope of the limit order book. A simple slope measure can be calculated by starting at the midpoint of the bid-ask spread (where the assumption that the quoted depth is zero) and rising to the cumulated depth somewhere within the limit order book.

It may well be that assuming slope is linear may mischaracterize the limit order book. Naes and Skjeltorp (2006) measure slope as an approximate elasticity at each price point then average the local slopes across all price levels. This sophisticated approach still uses an average, thereby reducing any variation to a single slope estimate.

In extant literature liquidity measures were often depicted as symmetric. For example, depth on each side of the bid-ask spread was often averaged to show the total depth in the market, irrespective of the side of the market. This implicit assumption does not take into consideration the different traders that may transact on each side of the market nor does it reflect the price pressure that may result from order imbalances. Michayluk and Neuhauser (2008) recognize the directional pressure and estimate a number of asymmetric liquidity measures. Their measure of depth asymmetry, called scaled depth difference, reflects the difference between the depth at the ask price and the depth at the bid price, all scaled by total depth.

$$ScaledDepthDifference = \frac{Q_A - Q_B}{Q_A + Q_B} \tag{L.4}$$

where Q_A is the depth of the ask quote and Q_B is the depth of the bid quote. Scaled depth difference reflects the measure of asymmetry and does not capture the overall level of depth therefore must be used in conjunction with depth measures.

3.3 Price impact

While not entirely a liquidity dimension on its own, price impact measures the subsequent price change following a transaction. Price impact may be the complement to resiliency component since it represents a change from the prior liquidity level: i.e., what is not resilient. The simple price impact for a trade that occurs at time t (stock subscripts suppressed) is:

$$SimplePriceImpact_t = 2q_t\left(m_{t+5\min} - m_t\right)/m_t \tag{L.5}$$

where q_t is a trade direction indicator (q=+1 for buyer-initiated trades and q=-1 for seller-initiated trades), and m_t is the midquote at time t. We use 5 minutes to determine the subsequent midpoint price but others including Bessembinder (2003) have used the prices 30 minutes later. If the trade occurs in the last five minutes of trading or last 30 minutes, depending on the subsequent trade chosen for the comparison, the final trade price of the day can be used.

Besides the mechanics of the calculation, there is a concern that the price change may be temporary or permanent. The average daily volume may provide one clue about the extent of trading and whether the trade size observed at time t is 'absorbed' by time t + 5 minutes (or 30 minutes). There may be temporary or 'transitory' price effects after a trade and the most appropriate time period to use is subjective. Any permanent change may reflect information but the longer the time period that is chosen for the subsequent comparison, the more likely that additional information may enter the market and confound the analysis. Conclusions about the price effect therefore must be considered in this context.

A common way to measure price impact at the stock-day level is by calculating the dollar volume-weighted average of the simple price impact for each trade within the stock-day. Alternatively, intraday windows can be determined to examine whether the price impact is stronger or weaker during times of higher or lower trading volume as well as time of the day.

Another consideration is the trade size and the positive relationship between trade size and price impact. Lin, Sanger and Booth (1995) identified the link between trade size and components of the bid ask spread. Specifically, the adverse selection component was shown to increase with trade size. Keim and Madhavan (1996) confirm that large block price impacts are a concave function of order size and a decreasing function of market capitalization. These findings suggest that with price impact, it may be prudent to consider the trade size or possibly even accumulated trades over a time window relative to market liquidity (possibly depth), when calculating price impact. Accumulation windows may be particularly important with the reductions in trade size in today's high frequency trading markets. Keim and Madhavan also suggest that if longer windows are used to assess price impact, an adjustment for market movements must be considered.

There is an accounting relation between effective spreads, realized spreads and simple price impacts that should also be considered:

$$SimplePriceImpact = EffectiveSpread - RealizedSpread \tag{L.6}$$

This relationship demonstrates the interconnected nature of these liquidity metrics and the difficulty in interpreting individual metrics without considering other measures.

There are also measures of price impact that specifically relate the price change to the extent of trading. Kyle's (1985) Lambda is one such measure and is actually a measure of illiquidity since a larger number indicates a less liquid market. It measures the price change over the dollar volume of trading. In the equation (L.7) the time period, *t*, can be set over any window.

$$\lambda = \frac{|\Delta \Pr ice_t|}{Volume_t} \quad \text{(L.7)}$$

A lower frequency (and less data intensive) way of calculating price impact is Amihud's (2002) illiquidity (*ILLIQ*) measure. This measure is calculated as the daily ratio of absolute stock return to its dollar volume, averaged over some period and it can be interpreted as the daily price response associated with one dollar of trading volume.

$$ILLIQ = \frac{\sum_1^T \frac{|return_t|}{dollarvolume_t}}{T} \quad \text{(L.8)}$$

where *t* indexes trading days and only positive volume trading days are included. Goyenko et al. (2009) extend the Amihud (2002) measure by separating the liquidity and the non-liquidity components of *ILLIQ*. They decompose the return and their new numerator only considers a proxy for the percent effective spread.

Pastor and Stambaugh (2003) develop a price impact measure, gamma, by running a regression of a stock's excess return (on day t + 1) on the prior day's market return and prior day's signed individual stock volume. The resultant estimate is therefore the reverse of the previous day's order flow shock. The value would be expected to be negative and the larger the absolute value the larger the price impact.

$$r_{t+1}^e = \vartheta + \phi r_t + (gamma) sign(r_t^e)(volume_t) + \varepsilon_t \quad \text{(L.9)}$$

where r_t^e is the stock's excess return above a value-weighted market index on day *t* and $volume_t$ is the dollar volume on day *t.*

The Pastor and Stambaugh (2003) gamma measure is really a measure of illiquidity and it could be argued that since this is the reversal it is really a measure of the temporary component. Price impact is therefore a much more complex concept than might be first anticipated. The Amihud (2002) and Pastor and Stambaugh (2003) measures can also be interpreted as inverse measures of disagreement observable in the market. Harris and Raviv (1993) suggest that all measures of price impact are affected by whether investors agree on the interpretation of arriving information since if there is agreement prices will move with little volume, which will mean low turnover and high price impact measures. Therefore, low price impact measures for low turnover firms may simply indicate that low turnover firms have low disagreement and are not necessarily less liquid.

All these complementary and conflicting interpretations of price impact measures provide an excellent opportunity for researchers to continue to refine the measures as well as allowing them to add many different interpretations to results when using the measures. The changing markets and extreme differences between the most liquid markets and the emerging markets and lower liquidity stocks continue to suggest that refinements and improvements are needed in this literature.

4. PRICE DISCOVERY

Price discovery, a fundamentally important role of financial markets, is the "efficient and timely incorporation of the information implicit in investor trading into market prices" (Lehmann, 2002, p. 259). The quality of price discovery has important effects on the real economy: better price discovery results in more informationally efficient prices, which in turn directs investment capital to its most productive uses and thus promotes macroeconomic growth.

Empirical market microstructure techniques can be used to gain insight into three key questions regarding price discovery:

- How good is aggregate price discovery?
- How does price discovery occur?
- Where does price discovery occur?

The next section (Section 4.1) will address the first of these questions by describing a series of measures of the aggregate quality of price discovery: the informational efficiency of prices. We deliberately present metrics that can all be calculated at a high frequency (calculated each stock-day using intraday data) so that they could be used to study the dynamics of informational efficiency: how it changes in response to different market structure or regulation changes, how it varies and co-varies through time and across stocks.

The section after that (Section 4.2) will address the second of these questions by describing a method of inferring the private information content of trades. This method can be used to gain insights into the process of how the market extracts information from order flow, what types of order flow are most informative and how the information contained in order flow changes in response to different market structures and regulatory changes, i.e., how the behavior of informed traders changes. We present the method in the classic setting of inferring the private information of the trade initiator but note that the technique easily generalizes to other informative events such as limit order arrivals, order cancellations, and so on.

Finally, Section 4.3 addresses the third key questions regarding price discovery by describing a set of techniques to quantify the contribution of different price series (markets, order flow sources, asset classes, price types) to price discovery. These techniques can be used to gain an understanding of where information first enters the market and makes prices informative, which is a function of where and how informed traders prefer to trade, among other things. In addition to presenting traditional metrics to achieve this goal, the section also describes new developments/metrics in the area (e.g., how to avoid the bias inherent in traditional metrics), which give rise to a number of new research opportunities in the area of price discovery.

4.1 High frequency informational efficiency metrics

In order to analyze the absolute amount of information that is impounded in prices we will estimate five informational efficiency metrics used in empirical studies: autocorrelation-based measures, variance ratios, short-term volatility, the delay in reflecting market-wide information, and the degree of short-term return predictability using lagged order imbalance. All of these metrics can be calculated at the stock-day frequency using intraday data.

4.1.1 Midquote return autocorrelations

Positive or negative midquote return autocorrelations suggest quotes deviate from a random walk process and have some short-term predictability, which is inconsistent with a highly efficient market. Unlike trade prices, midquotes do not suffer from bid-ask bounce and are often interpreted as a market's best estimate of the fundamental value. We can calculate first-order return autocorrelations for each stock-day, at various intraday frequencies, $k \in \{10\,\mathrm{sec},\ 30\,\mathrm{sec},\ 60\,\mathrm{sec}\}$, similar to Hendershott and Jones (2005):

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

where $r_{k,t}$ is the t^{th} 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 over-reaction of returns to information, with larger values indicating greater inefficiency (e.g., Boehmer and Wu, 2013).

We can also compute a combined autocorrelation measure, $Autocorrelation_{Factor}$, by taking the first principal component of the absolute autocorrelations at the three frequencies. The purpose of calculating the first principle component is to identify inefficiency from the common variance in the absolute autocorrelations at difference frequencies.

4.1.2 Variance ratios

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 walk (e.g., Lo and MacKinlay, 1988). We can calculate three variance ratios for each stock-day at different intra-day frequencies, similar to O'Hara and Ye (2011):

$$VarianceRatio_{kl} = \left| \frac{\sigma^2_{kl}}{k\sigma^2_l} - 1 \right| \tag{2}$$

where σ^2_l and σ^2_{kl} 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 can also compute a combined variance ratio, $VarianceRatio_{Factor}$, by taking the first principal component of the three variance ratios. Again, the purpose of calculating the first principle component is to identify inefficiency from the common variation in the variance ratios at difference frequencies.

4.1.3 Short-term volatility

For each stock-day we can also combine the intra-day midquote return standard deviations calculated at 10, 30 and 60 second frequencies by taking their first principal component. This produces a single measure of short-term volatility, which is a proxy for noise and temporary deviations of prices from their equilibrium values due to trading frictions (e.g., O'Hara and Ye, 2011).

4.1.4 Delay in impounding public information

One measure of short-term return predictability 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 can estimate a regression of 1-minute midquote returns for stock i, $r_{i,t}$, on the index return, $r_{m,t}$, and ten lags (suppressing day subscripts):

$$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \sum_{k=1}^{10} \delta_{i,k} r_{m,t-k} + \varepsilon_{it} \quad (3)$$

We save the R^2 from the above unconstrained regression, $\mathrm{R}^2_{Unconstrained}$, 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, $\mathrm{R}^2_{Constraind}$. *Delay* is then calculated as:

$$Delay = 1 - \frac{\mathrm{R}^2_{Constraind}}{\mathrm{R}^2_{Unconstrained}} \quad (4)$$

and takes values between 0 and 1. 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.

4.1.5 Predictability of midquote returns based on lagged order imbalance

Another measure of short-term return predictability is the extent to which midquote returns can be predicted using lagged order imbalances, similar to Chordia et al. (2005, 2008). For each stock-day we can estimate the following regression of 1-minute midquote returns, r_t, on ten lags of 1-minute order imbalance (buyer-initiated less seller-initiated dollar volume of trades), OIB_t (suppressing stock-day subscripts):

$$r_t = \alpha + \sum_{k=1}^{10} \beta_k OIB_{t-k} + \varepsilon_t \quad (5)$$

Following Chordia et al. (2008) the R^2s of the regression are a proxy for inefficiency, $Inefficiency_{OIB}$.

4.2 Measuring the information content of a trade

A simple measure of the private information contained in a trade is the amount by which it changes the midquote (often viewed as the market's best estimate of the fundamental value) once the market has had sufficient time to react to the trade. For example, one might calculate a five-minute simple price impact as the change in the midquote from immediately prior to the trade to five minutes after the trade. If trades contain information (strictly speaking, trade initiation contains information) then midquotes should be revised up following buyer-initiated trades and revised down following seller-initiated trades. This notion is empirically supported.

To be able to aggregate simple price impacts across trades and not cancel out the positive (average) price impacts of buyer-initiated trades with the negative (average) price impacts of seller-initiated trades, the simple price impact is often defined with a trade direction indicator (q = +1 for buyer-initiated trades and q = –1 for seller-initiated trades):

$$SimplePriceImpact_t = q_t\left(m_{t+5\min} - m_t\right)/m_t \tag{6}$$

where m_t is the midquote at time t. Simple price impacts are sometimes defined as above but multiplied by a factor of two so that they are easily compared to effective and realized spreads via the accounting relation: *EffectiveSpread* = *RealizedSpread* + *SimplePriceImpact*.

Simple price impacts, however, are misleading when trades are serially correlated, which is the case in most market settings. This is because the price impact of other trades that occur between time t and t + 5min impacts the measurement of the price impact of the trade at time t. For example, suppose 10 buyer-initiated trades occur in quick succession and each causes the midquote to be revised upward by 1 basis point. The first trade will be credited with a price impact of 10 bps, although its true price impact was only 1 bp. Thus the simple price impact will tend to overstate the informativeness of trades, and if it is used to compare the informativeness of two or more types of trades or sources of order flow it will be biased towards the trade type or order flow source that has stronger serial correlation.

To address this problem (and other reasons), Hasbrouck (1991) proposes measuring the information content of trades using vector auto-regression (VAR) models of order flow and returns. Many variants of this technique have been used in the literature, but all generally involve systems of equations modeling signed order flow (measured as volume, dollar volume, or trades) and log returns (e.g., midquote returns to avoid bid-ask bounce).

A simple example of such VARs, in its structural form, is as follows:

$$\begin{aligned} x_t &= \mu^x + \sum_{i=1}^{60}\phi_i^r r_{t-i} + \sum_{i=1}^{60}\phi_i^x x_{t-i} + \varepsilon_t^x \\ r_t &= \mu^r + \sum_{i=1}^{60}\theta_i^r r_{t-i} + \sum_{i=0}^{60}\theta_i^x x_{t-i} + \varepsilon_t^r \end{aligned} \tag{7}$$

where t indexes 1-second intervals (individual stock and date subscripts are suppressed), x_t is signed dollar volume of trades in the 1-second interval, t, r_t is the log-midquote change in the t^{th} interval, ε_t^x is unanticipated signed volume, and ε_t^r is a midquote innovation not caused by order flow. Each equation (midquote returns and signed dollar volume) contains 60 lags of each variable. The number of lags can be chosen using information criteria (e.g., Akaike Information Criterion) or simply by economic reasoning of consideration for the speed and dynamics of the market. In addition to the 60 lags, midquote returns are also determined by contemporaneous order flow, but not the other way around (notice the summation begins at $i = 0$ for lags of signed dollar volume in the return equation, but $i = 1$ for lags of returns in the signed dollar volume equation).

The assumption about contemporaneous causality follows from economic reasoning that if the VAR is estimated at sufficiently high frequency trades cannot react to midquote changes instantly (in the same sampling period), but trades can cause contemporaneous changes in the midquote (e.g., by executing all the resting limit orders at the best price step). The contemporaneous causality assumptions are important for interpreting the results from the VAR model, and are what makes the above model the structural form.

One novelty of this approach is the recognition that information is conveyed through trade *innovations* – the unexpected portion of order flow – not by total transaction volume, which has a considerable predictable component due to serial correlation in volume. Furthermore, the approach allows for separation of midquote innovations into those that are driven by trade innovations (private information held by the trade initiator) and those not associated with trade innovations (public information being reflected in quotes by liquidity providers). This separation is an important part of identifying the private information content of a trade.

Hasbrouck (1991) proposes that the information content of a trade can be meaningfully measured as the trade innovation's ultimate (permanent) impact on the midquote. This is operationalized using the impulse response function, which measures the response of the system (or any of the variables) to an unanticipated shock to one of the variable while the system is in a steady state (all past lagged values are zero). Hasbrouck (1991, p.187) provides a simple example of how to "step forward" an estimated VAR system and measure the cumulative midquote revision in response to an unanticipated buyer-initiated trade:

$$r_t = 1.998\ x_t + .096\ x_{t-1} + .053\ x_{t-2} + .025\ x_{t-3} + \hat{v}_{1,t}$$

$$x_t = -.486\ x_{t-1} - .269\ x_{t-2} - .124\ x_{t-3} + \hat{v}_{2,t}$$

where r_t is the quote-revision, x_t is the trade, $\hat{v}_{1,t}$ is the quote-revision innovation, $\hat{v}_{2,t}$ is the trade innovation, and t indexes transactions. Based on an initial innovation at transaction $t = 0$ of $\hat{v}_{1,t} = 0$ and $\hat{v}_{2,t} = 1$ (i.e., a purchase of one share), the table gives the implied response of r_t and x_t. Quantity α_m is the cumulative quote revision through transaction m: $\alpha_m = \sum_{i=0}^{m} r_i$.

t	r_t	x_t	α_m
0	1.998	1.0	1.998
1	−.874	−.486	1.124
2	−.060	−.033	1.064
3	.041	.023	1.105
4	.105	.058	1.210
5	−.054	−.030	1.155
⋮			
10	-6×10^{-4}	-3×10^{-4}	1.156

Initially, all lagged values are zero. At time t = 0 the trade innovation is shocked with one unanticipated buyer-initiated trade for one share and the cumulative response of the midquote, a_m, is measured at every step. Notice that the return and trade responses get weaker with every step and would eventually approach zero (in stable systems). Taking t = 10 as the “limiting point of the system” (the point at which the system has fully incorporated the effects of the unanticipated trade at t = 0), Hasbrouck interprets the a_m = 1.156 as the information content of the trade. Although in theory the ultimate (permanent) impact of the trade is achieved only after the system is stepped forward an infinite number of time steps, in practice the permanent or ultimate impacts are often measured after stepping the system through a finite sufficient number of steps so that *most* of the impact is felt.

This example illustrates another advantage of the VAR approach to measuring price impact. The midquote initially overshoots in response to the trade innovation, which could occur due to inventory management by liquidity providers (raise quotes in response to a buy trade *above* the revised expectation of fundamental value in order to attract sellers to offset the inventory position incurred from accommodating the buy trade). As the VAR is stepped forward in time the model predicts offsetting seller-initiated order flow and a subsequent reversal of the initial midquote overshoot. Therefore, in estimating the information content of a trade, the VAR is able to accommodate over/undershooting of initial reactions to trades as well as slow adjustment and various transient effects.

Turning to the practical aspects of estimating and interpreting microstructure VAR models, through substitution the structural VAR model given in equation (7) can be re-expressed in reduced form (i.e., with only lagged (pre-determined) coefficients on the right hand side),

$$\begin{aligned} x_t &= u^x + \sum_{i=1}^{60} \alpha_i^r r_{t-i} + \sum_{i=1}^{60} \alpha_i^x x_{t-i} + e_t^x \\ r_t &= u^r + \sum_{i=1}^{60} \beta_i^r r_{t-i} + \sum_{i=1}^{60} \beta_i^x x_{t-i} + e_t^r \end{aligned} \tag{8}$$

where the reduced form parameters, u, α, and β, and error terms, e, are composites of the structural model parameters and error terms.[4] Interestingly (and sometimes confusingly) many different structural VARs, with different assumptions about contemporaneous causality, collapse down to the same reduced form model. The reduced form errors in the VAR model above are $e_t^x = \varepsilon_t^x$ and $e_t^r = b_1\varepsilon_t^x + \varepsilon_t^r = b_1 e_t^x + \varepsilon_t^r$. Therefore, under the assumption that the structural errors are serially uncorrelated and i.i.d. (and uncorrelated across equations), the reduced form errors are also serially uncorrelated and i.i.d., but are contemporaneously correlated across equations (related linearly by b_1). The variance-covariance matrix of the reduced form errors is:

$$\Omega = \begin{bmatrix} \sigma_{e^x}^2 & \sigma_{e^x e^r} \\ \sigma_{e^x e^r} & \sigma_{e^r}^2 \end{bmatrix} \tag{9}$$

Often, for econometric reasons the reduced form VAR is estimated and then the reduced form estimates are used to infer the dynamics of the structural model. The reduced form VAR can be estimated equation-by-equation using OLS or if the variables on the right hand side differ across the equations then there may be

---

[4] For example, consider a simple structural VAR:

$$\begin{aligned} x_t &= \alpha_1 r_{t-1} + \alpha_2 x_{t-1} + \varepsilon_{1,t} \\ r_t &= \alpha_3 r_{t-1} + \alpha_4 x_t + \alpha_5 x_{t-1} + \varepsilon_{2,t} \end{aligned}$$

Substituting the first equation into the second we get:

$$r_t = \alpha_3 r_{t-1} + \alpha_4\left(\alpha_1 r_{t-1} + \alpha_2 x_{t-1} + \varepsilon_{1,t}\right) + \alpha_5 x_{t-1} + \varepsilon_{2,t}$$

Rearranging terms

$$r_t = \left(\alpha_3 + \alpha_4\alpha_1\right) r_{t-1} + \left(\alpha_4\alpha_2 + \alpha_5\right) x_{t-1} + \left(\varepsilon_{2,t} + \alpha_4\varepsilon_{1,t}\right)$$

we obtain the reduced form:

$$\begin{aligned} x_t &= b_1 r_{t-1} + b_2 x_{t-1} + e_{1,t} \\ r_t &= b_3 r_{t-1} + b_4 x_{t-1} + e_{2,t} \end{aligned}$$

where $b_3 = \left(\alpha_3 + \alpha_4\alpha_1\right)$, $b_4 = \left(\alpha_4\alpha_2 + \alpha_5\right)$, and $e_{2,t} = \left(\varepsilon_{2,t} + \alpha_4\varepsilon_{1,t}\right)$.

efficiency gains from using Seemingly Unrelated Regression (SUR). The reduced form parameter estimates describe the dependence of the variables on the past, whereas the variance-covariance matrix contains information on the contemporaneous effects. Therefore, it does not make sense to examine a reduced form shock of say $e_0^x = 1,\ e_0^r = 0$ unless the reduced form VAR *is* the structural VAR we have in mind (i.e., we believe there are no contemporaneous relations), because we do not know what the structural interpretation of the shock is – that depends on what structural model we have in mind.

Rather, we can look at the structural model and come up with a structural shock of interest, say $\varepsilon_0^x = B,\ \varepsilon_0^r = 0$ corresponding to $*B* of unanticipated buyer-initiated trading. We can then infer the equivalent reduced form shock, feed it through the estimated reduced form VAR and examine the dynamic response of the variables of interest.[5] In this example we would examine the response to the equivalent reduced form shock $e_0^x = B,\ e_0^r = \frac{\sigma_{e^x e^r}}{\sigma_{e^x}^2} B$.[6] In contrast, a structural shock to the midquote (which could arise due to public information entering in the market and causing a revision in quotes without any associated trading) of magnitude *R*, $\varepsilon_0^x = 0,\ \varepsilon_0^r = R$, has a similar equivalent reduced form shock of $e_0^x = 0,\ e_0^r = R$ because of the absence of contemporaneous causality from returns to volume in the structural model by assumption (based on economic arguments). In computing the impulse response we set the intercepts to zero and all past lagged values to zero, so that the impulse responses can be interpreted as relative to the dynamics of the system with no shock.

Many interesting extensions and modifications are possible to the simple VAR example discussed above, including: (i) estimating the model in trade time rather than clock time (subscript *t* could index trade events), which is well suited to infrequent and irregularly spaced trades; (ii) inclusion of other variables in the system, such as the spread or limit order arrivals; (iii) adding non-linearity, e.g., squared volume

[5] In computing the response to the equivalent reduced form shock we would use the reduced form parameter estimates because we are using a reduced form shock. The alternative is to back out the structural parameters from the reduced form parameters and then examine the shock to the structural model, in which case we would use the structural shock with the structural parameters (not the equivalent reduced form shock).

[6] Typical shock magnitudes are unit (1) or one standard deviation of the reduced form errors for the corresponding equation, e.g., for signed dollar volume σ_{e^x}.

terms; and (iv) extension to multiple types or sources of order flow. Some example of different order flow types/sources include different traders (e.g., local versus foreign, institution versus retail, HFT versus non-HFT), different markets (e.g., different trading venues for cross-listed stocks, on-market versus off-market trades, or different stock exchanges trading different stocks), different order types (e.g., short sales versus long sales, limit versus market orders), and different size orders.

An example of extending the two-equation VAR above to one with two sources of trades, the primary market and the other market for cross-listed stocks, is as follows:

$$\begin{aligned}
x_t^{Primary} &= \mu^{Primary} + \sum_{i=1}^{60} \phi_i^{Primary} x_{t-i}^{Primary} + \sum_{i=1}^{60} \phi_i^{Other} x_{t-i}^{Other} + \sum_{i=1}^{60} \phi_i^{r} r_{t-i} + \varepsilon_t^{Primary} \\
x_t^{Other} &= \mu^{Other} + \sum_{i=1}^{60} \theta_i^{Primary} x_{t-i}^{Primary} + \sum_{i=1}^{60} \theta_i^{Other} x_{t-i}^{Other} + \sum_{i=1}^{60} \theta_i^{r} r_{t-i} + \varepsilon_t^{Other} \\
r_t &= \mu^{r} + \sum_{i=0}^{60} \gamma_i^{Primary} x_{t-i}^{Primary} + \sum_{i=0}^{60} \gamma_i^{Other} x_{t-i}^{Other} + \sum_{i=1}^{60} \gamma_i^{r} r_{t-i} + \varepsilon_t^{r}
\end{aligned} \tag{10}$$

The contemporaneous causality assumptions in the structural VAR above are that signed trading volume on either market can contemporaneously affect the midquote (average of the midquotes of the two trading venues) but not the other way around, and neither of the volume types can contemporaneously cause the other volume type. The reduced form errors in the VAR model above are $e_t^{Primary} = \varepsilon_t^{Primary}$, $e_t^{Other} = \varepsilon_t^{Other}$ and $e_t^{r} = b_1\varepsilon_t^{Primary} + b_2\varepsilon_t^{Other} + \varepsilon_t^{r} = b_1 e_t^{Primary} + b_2 e_t^{Other} + \varepsilon_t^{r}$. Therefore, reduced form shock that is equivalent to a one-unit structural shock to one of the volume types, say the primary market (structural shock $\varepsilon_0^{Primary} = 1,\ \varepsilon_0^{Other} = 0,\ \varepsilon_0^{r} = 0$), is $e_0^{Primary} = 1,\ e_0^{Other} = 0,\ e_0^{r} = \frac{\sigma_{e^{Primary}e^{r}}}{\sigma^2_{e^{Primary}}} 1$. The reduced form shock that is equivalent to a one-unit structural shock to midquote returns (structural shock $\varepsilon_0^{Primary} = 0,\ \varepsilon_0^{Other} = 0,\ \varepsilon_0^{r} = 1$), is $e_0^{Primary} = 0,\ e_0^{Other} = 0,\ e_0^{r} = 1$.

In the absence of theory that guides the assumptions about contemporaneous relations a common approach is to use Cholesky factorization (or Cholesky decomposition) of the variance-covariance matrix and calculate orthogonalized impulse response functions. We won't go into the details here other than to say that this method corresponds to a structural model with recursive causality running from the first variable down the list of variables, i.e., the first variable can contemporaneously cause all of the other variables, the second variable can contemporaneously cause all of the other variables except for the first, and so on until the last variable which can be contemporaneously affected by any of the other variables but cannot contemporaneously affect any of the other variables. For example, in our earlier three-equation VAR the orthogonalized impulse response functions using Cholesky factorization corresponds to a structural model as follows:

$$\begin{aligned}
x_t^{Primary} &= \mu^{Primary} + \sum_{i=1}^{60} \phi_i^{Primary} x_{t-i}^{Primary} + \sum_{i=1}^{60} \phi_i^{Other} x_{t-i}^{Other} + \sum_{i=1}^{60} \phi_i^{r} r_{t-i} + \varepsilon_t^{Primary} \\
x_t^{Other} &= \mu^{Other} + \sum_{i=0}^{60} \theta_i^{Primary} x_{t-i}^{Primary} + \sum_{i=1}^{60} \theta_i^{Other} x_{t-i}^{Other} + \sum_{i=1}^{60} \theta_i^{r} r_{t-i} + \varepsilon_t^{Other} \\
r_t &= \mu^{r} + \sum_{i=0}^{60} \gamma_i^{Primary} x_{t-i}^{Primary} + \sum_{i=0}^{60} \gamma_i^{Other} x_{t-i}^{Other} + \sum_{i=1}^{60} \gamma_i^{r} r_{t-i} + \varepsilon_t^{r}
\end{aligned} \tag{11}$$

When using this approach the order of the variables in the system matters. One way to test the robustness to different orderings is to estimate the orthogonalized impulse response to a variable when it is the first variable in the system and then to repeat the procedure with the variable as the last in the system – the two sets of impulse responses can be loosely interpreted as upper and lower bounds.

Finally, one other approach to calculating impulse responses is the "generalised impulse response function" of Pesaran and Shin (1998). This method is like using orthogonalized impulse responses with the novel feature that when computing the responses to a particular variable that variable is always placed as the first variable in the recursive system.

4.3 Assigning contributions to price discovery among multiple prices

4.3.1 The economic process of price discovery

Price discovery, a fundamentally important role of secondary markets, is the "efficient and timely incorporation of the information implicit in investor trading into market prices" (Lehmann, 2002, p. 259). When multiple price series are related via a common asset (e.g., a stock trading in multiple venues, order flow for one security from different market participants, and different derivative securities linked to the same underlying asset) the contribution of a price series to price discovery is typically considered to be the extent to which it is the first to reflect new information about the 'true' underlying asset value. In order to correctly interpret the price discovery metrics it is essential to distinguish between speed and noise in the process of price discovery. Both are implicit in Lehman's (2002) definition of price discovery provided earlier, "*efficient* and *timely* incorporation of information …". *Timely* refers to the relative speed with which a price series reflects new information about the fundamental value. *Efficient* implies a relative absence of noise, such as bid-ask bounce, tick discreteness, temporary deviations due to imperfect liquidity and so on.

In a multiple price series setting, by far the most widely accepted view of what it means for a price series to contribute to or to be dominant in price discovery stems from the seminal work of Hasbrouck (1995), which puts forward the "who moves first" view of price discovery. Hasbrouck proposes a measure of price discovery, the 'information share' (*IS*), intended to measure "'who moves first' in the process of price adjustment" (p.1184) to innovations in the efficient price. Hasbrouck is very clear on what *IS* does not measure; it does not measure the total amount of information impounded into prices, nor does it measure which market has the "best" prices. Hasbrouck illustrates the latter with an example of a market that is informationally dominant (and has an *IS* of 100%) because innovations in the market drive reactions in other markets, yet the informationally dominant market could also have the widest spreads and therefore not necessarily the "best" prices.

The focus in the price discovery literature on which price series moves first rather than which is most informative follows from the objective of uncovering where and how information enters a market and gets impounded into prices. Knowing which price "moves first" regardless of how noisy it is contributes to achieving this objective; knowing which price is more informative does not necessarily.

4.3.2 The empirical measures of price discovery[7]

Turning to the empirical measures of price discovery, traditionally there have been two main (occasionally competing) empirical measures of the contribution of different price series to price discovery: Hasbrouck information share (*IS*) and Gonzalo-Granger component share (*CS*). Both *IS* and *CS* rely on the notion that prices for the same asset (in different markets, for example) can deviate from one another in the short run due to trading frictions, but will converge in the long run because both are connected to the fundamental value of the asset. Such prices series

[7] For a good review of *IS* and *CS* and how they are related see Baillie et al. (2002), Lehman (2002) and Yan and Zivot (2010). This section provides a brief overview of these metrics.

are therefore cointegrated and lend themselves to empirical analysis using vector error correction models (VECM) under the assumption that the price series share a common random walk efficient price. Both *IS* and *CS* are derived from the estimates of a reduced form VECM.

Hasbrouck (1995) proposes that the contribution of a price series to price discovery (the 'information share', *IS*) can be measured by the proportion of the variance in the common efficient price innovations that is explained by innovations in that price series. This approach follows naturally as an extension of Hasbrouck's earlier work (Hasbrouck, 1991) in which he proposes that the relative informativeness of trades could be measured by the proportion of efficient price variation attributable to trades. When price innovations across markets are correlated, the attribution of efficient price innovation variance cannot be done uniquely and instead one can estimate an upper and lower bound on a price series' *IS*.

The second commonly used metric, *CS*, is based on Gonzalo and Granger's (1995) work on the econometrics of cointegration, and is first applied to price discovery by Booth et al. (1999), Chu et al. (1999) and Harris et al. (2002). Gonzalo and Granger (1995) propose a method of decomposing a cointegrated price series into a permanent component and a temporary component using the error correction coefficients. In the context of price discovery the permanent component is interpreted as the common efficient price and the temporary component reflects deviations due to trading fractions. Importantly, Gonzalo and Granger (1995) show that the permanent component is a linear combination of all variables in the cointegration system (all the price series). Booth et al. (1999), Chu et al. (1999) and Harris et al. (2002) propose that a price series with greater weight in the linear combination moves more closely with the common efficient price and thus contributes more to price discovery. Therefore, under this approach a price series' contribution to price discovery (the 'component share', *CS*) is it's normalized weight in the linear combination of prices that forms the common efficient price. Baillie et al. (2002) show that although *IS* and *CS* seem dissimilar, they share a lot in common because both are closely related to the same combination of the reduced form error correction coefficients.

Lehmann (2002) points out that because *IS* and *CS* are both defined in terms of a reduced form VECM, their interpretation with respect to price discovery is not always clear because it is dependent on the (often unspecified) structural model of price formation. Yan and Zivot (2010) directly address this problem by specifying a fairly general structural cointegration model for asset prices and analytically demonstrating what *IS* and *CS* measure for that model. Their structural model, motivated by models used in empirical macroeconomics, consists of two price series that are driven by two sources of shocks: one permanent and one transitory. The permanent shocks represent innovations in the fundamental value and therefore, by definition, a one-unit permanent shock leads to a one-unit increase in each of the prices in the long run. The transitory shock represents noise due to trading frictions and therefore, by definition, a one-unit transitory shock has zero effect on the prices in the long run. The short-run impacts of the permanent and transitory shocks, however, are not restricted by the structural model; each price series is defined by two lag polynomials that describe its dynamic response to the permanent and transitory shocks. This allows the price series to differ in how quickly they reflect innovations in the fundamental value and how they are impacted by the transitory shocks.

Within this setup, Yan and Zivot (2010) are able to express the *IS* and *CS* of a price series in terms of its dynamic response to permanent and transitory shocks (the lag polynomials). Their results indicate that *CS* is a function of the dynamic responses of the two price series to the transitory shocks only, whereas *IS* is a function of the dynamic responses of the two price series to the transitory *and* permanent shocks. This suggests that *IS* and *CS* can give misleading information regarding price discovery in some situations due to their dependence on the dynamic response to transitory shocks. A useful byproduct of this result is that *IS* and *CS* can be combined in an expression, $|(IS_1/IS_2)(CS_2/CS_1)|$, such that the dynamic responses to the transitory shocks cancel out, leaving only a ratio of the price series' dynamic responses to the permanent shocks. This metric can be useful in attributing price discovery because under certain assumptions it accurately measures the relative impact of permanent shocks on the two price series and, unlike *IS* and *CS*, it is not influenced by how the price series respond to transitory shocks. The key assumptions underlying Yan and Zivot's (2010) results are: (i) there are only two price series; (ii) the structural model has only one permanent and one transitory shock; and (iii) the reduced form VECM errors are uncorrelated. For models of price formation that violate these assumptions it is not known to what extent the expression that combines *IS* and *CS* purges the responses to transitory shocks.

Putnins (2013) uses the analytic results of Yan and Zivot (2010) to construct a new price discovery metric, the "information leadership share" (*ILS*), which is comparable to the *IS* and *CS* because it takes values in the range [0,1]. Putnins (2013) shows that unlike *IS* and *CS*, the *ILS* is not affected by differences in noise levels in the two price series – *ILS* simply measures which price series is the first to reflect innovations in the fundamental value, the "who moves first" aspect of price discovery.

Yan and Zivot (2010) and Putnins (2013) show that both of the traditional price discovery metrics, *IS* and *CS*, are biased in the presence of different levels of microstructure noise, explaining why they can give conflicting signals about price discovery. Putnins (2013) shows that this bias is so substantial that it causes many existing studies to reach conclusions that are polar opposites of what an unbiased metric would suggest. For example, Rittler (2012) concludes that futures dominate price discovery in the EU emission trading market and that their dominance has increased through time, whereas *ILS* indicates the opposite: the spot market leads price discovery overall, and the contribution of the futures market to price discovery has declined through time. Similarly, Hsieh et al. (2008) conclude that futures dominate options in price discovery for the Taiwan stock index, whereas *ILS* suggests that the opposite is true. Mizrach and Neely (2008) conclude that the Treasury futures market plays an important role in price discovery in the US Treasuries market compared to the spot market, whereas *ILS* indicates the opposite is true. Finally, Chen and Gau (2010) conclude that the spot market provides more price discovery in currencies than does the future market, but *ILS* reveals the opposite is true. The choice of price discovery metric and correct interpretation of what the metric actually measures in the particular setting are crucial to drawing correct conclusions about where/how price discovery occurs. As these stark differences in conclusion suggest, there is considerable scope for future research to re-investigate a number of previously studied price discovery settings, including the role of different asset classes, market structures, types of traders, and types of prices.

4.3.3 Estimating the price discovery share metrics

For each stock-day we can estimate the following VECM on a grid of 1-second clock time intervals:

$$\begin{aligned}\Delta p_{1,t} &= \alpha_1(p_{1,t-1} - p_{2,t-1} - \mu) + \sum_{i=1}^{60}\gamma_i \Delta p_{1,t-i} + \sum_{j=1}^{60}\delta_j \Delta p_{2,t-j} + \varepsilon_{1,t} \\ \Delta p_{2,t} &= \alpha_2(p_{1,t-1} - p_{2,t-1} - \mu) + \sum_{k=1}^{60}\phi_k \Delta p_{1,t-k} + \sum_{m=1}^{60}\varphi_m \Delta p_{2,t-m} + \varepsilon_{2,t}\end{aligned} \quad . \qquad (12)$$

The price series, p_1 and p_2 could be any two cointegrated price series that are interesting to compare. For example, trade prices of different traders (e.g., local versus foreign, institution versus retail, HFT versus non-HFT), prices from different markets trading cross-listed stocks, on-market versus off-market trades, different order types (e.g., short sales versus long sales, limit versus market orders). In the empirical setting studied in this course, p_1 and p_2 correspond to midquotes from the primary and the other exchange, for a cross-listed stock.

The cointegrating vector in the VECM above (+1,-1), which defines the error correction term is obtained from theory and corresponds to settings where two prices are related to the same underlying asset and therefore have a 1:1 long-run equilibrium relation, enforced by arbitrage. In some settings a more general cointegrating vector may be required. Similar to the VAR models discussed in the previous section, the VECM above can also be estimated in clock time or event time (e.g., trade time, treating trades as the events that define increments of time, or order arrival time, etc.).[8]

We can calculate IS_1, IS_2, CS_1 and CS_2 from the error correction parameters and variance-covariance of the error terms, following Baillie et al. (2002). The component shares are obtained from the normalized orthogonal to the vector of error correction coefficients, $\alpha_\perp = (\gamma_1, \gamma_2)'$, thus:

$$CS_1 = \gamma_1 = \frac{\alpha_2}{\alpha_2 - \alpha_1}, \quad CS_2 = \gamma_2 = \frac{\alpha_1}{\alpha_1 - \alpha_2}. \tag{13}$$

Given the covariance matrix of the reduced form VECM error terms,

$$\Omega = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{14}$$

and its Cholesky factorization, $\Omega = MM'$, where

$$M = \begin{pmatrix} m_{11} & 0 \\ m_{12} & m_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2(1-\rho^2)^{1/2} \end{pmatrix}, \tag{15}$$

[8] Using 1-second clock time observations is the most common approach (e.g., Hasbrouck, 1995; Chakravarty et al., 2004; Kurov and Lasser, 2004; Anand and Subrahmanyam, 2008; Shastri et al., 2008). Frequent sampling is desirable as it tends to result in low contemporaneous innovations (correlation of reduced form errors) and thus narrow upper/lower bounds of Hasbrouck information shares. If one uses transaction time it is important to address the problem of non-syncronous trading, i.e., that at any point in time not all price series will have an updated value. Harris et al. (1995, 2002) use various data thinning algorithms (such as 'MINISPAN') to address the problem of non-synchronous trading. However, such data thinning can lead to incorrect inferences (Hasbrouck, 2002, 2003; Lehman, 2002).

the IS are calculated using:

$$IS_1 = \frac{(\gamma_1 m_{11} + \gamma_2 m_{12})^2}{(\gamma_1 m_{11} + \gamma_2 m_{12})^2 + (\gamma_2 m_{22})^2}, \qquad IS_2 = \frac{(\gamma_2 m_{22})^2}{(\gamma_1 m_{11} + \gamma_2 m_{12})^2 + (\gamma_2 m_{22})^2}. \quad (16)$$

Because the estimates of IS are affected by the ordering of the price series in the VECM (similar to the way the price impacts using orthogonalized impulse response functions are sensitive to the order of variables in the VAR), we can use the approach advocated by Baillie et al. (2002) (also used by Booth et al. (2002), Cao et al. (2009), Chen and Gau (2010), Korczak and Phylaktis (2010) and others) and calculate IS under each of the two possible orderings and then take the simple average.

We can calculate the Yan-Zivot information leadership metric using:

$$IL_1 = \left| \frac{IS_1}{IS_2} \frac{CS_2}{CS_1} \right|, \qquad IL_2 = \left| \frac{IS_2}{IS_1} \frac{CS_1}{CS_2} \right|. \quad (17)$$

Unlike IS and CS, the IL proposed by Yan and Zivot is not a "share" in that IL_1 and IL_2 do not sum to 1 (or even approximately 1). Rather, IL_1 has the range $[0,\infty)$. Values of IL_1 above (below) 1 suggest that p_1 leads (does not lead) the process of incorporating new information about the fundamental value. In order to make the information leadership metric easier to interpret and more readily comparable to IS and CS we can calculate the "information leadership shares" (ILS) of p_1 and p_2 as proposed by Putnins (2013):

$$ILS_1 = \frac{IL_1}{IL_1 + IL_2}, \qquad ILS_2 = \frac{IL_2}{IL_1 + IL_2}. \quad (18)$$

Now, ILS_1 and ILS_2 have the range [0,1], similar to IS and CS, with values above (below) 0.5 indicating the price series leads (does not lead) the process of adjusting to innovations in the fundamental value. In addition to being more readily comparable to IS and CS, ILS has the added advantage of not producing extreme values which occur in IL when both IS and CS approach the value 1.

4.3.4 Some applications of price discovery share metrics

The following is a list of various applications of price discovery metrics:

- Studies that compare the contributions to price discovery of different stock exchanges or trading venues that have similar structure using cross-listed stocks (e.g., Hasbrouck, 1995; Huang, 2002; Harris et al., 2002; Eun and

Sabherwal, 2003; Pascual et al., 2006; Frijns et al., 2010; Korczak and Phylaktis, 2010; Chen and Sub Choi, 2012; Chen et al., 2013).

- Studies that analyze the price discovery contributions of stock options compared to stocks (e.g., Chakravarty et al., 2004; Muravyev et al., 2013).
- Studies that analyze the price discovery contributions of futures compared to spot markets (e.g., Booth et al., 1999; Chu et al., 1999; Covrig et al., 2004; Shastri et al., 2008; Mizrach and Neely, 2008; Cabrera et al., 2009; Chen and Gau, 2010; Liu and An, 2011; and Rittler, 2012).
- Studies that analyze the price discovery contributions of options compared to futures (e.g., Booth et al., 1999; Hsieh et al., 2008; and Chen and Chung, 2012),
- Studies that analyze the price discovery contributions of CDS compared to bonds and stocks (e.g., Forte and Pena, 2009).
- Studies that analyze the price discovery contributions of screen-based compared to open-outcry (floor) trading (e.g., Ates and Wang, 2005).
- Studies that analyze the price discovery contributions of different contract maturities (e.g., Fricke and Menkhoff, 2011).
- Studies that analyze the price discovery contributions of quotes compared to trade prices (e.g., Cao et al., 2009).
- Studies that analyze the price discovery contributions of order flow from different types of traders (e.g., Fong and Zurbruegg, 2003; Kurov and Lasser, 2004; Anand and Subrahmanyam, 2008; and Anand et al., 2011).

5. ASSESSING THE IMPACT OF A MARKET STRUCTURE CHANGE

The empirical setting studied in this course (the switch to post-trade anonymity for a subset of Scandinavian stocks) provides a nice natural experiment in which we have observations on a set of stocks before and after the "treatment" (post-trade anonymity) is applied and observations on a set of stocks during the same periods of time that are not subject to the treatment. Such settings lend themselves naturally to analysis in a difference-in-differences framework. The attractive features of difference-in-differences approach are the ability to use very general controls for changes in the outcome variables through time that are not related to the treatment and controls for differences in the cross-sectional units, in our case stocks. We could estimate the following regression model:

$$y_{it} = \mu_i + \mu_t + \beta\left(D_i^{AnonStock} D_t^{AnonPeriod}\right) + \varepsilon_{it} \tag{19}$$

where y_{it} is any of the liquidity or price discovery metrics described in these notes (or any other metric) calculated for each stock-day *it*, μ_i and μ_t are stock and time fixed effects, respectively, $D_i^{AnonStock}$ is a dummy variable that takes the value of 1 for stocks that switched to post-trade anonymity and zero otherwise, and $D_t^{AnonPeriod}$ is a dummy variable that takes the value of 1 during the period in which some stocks had post-trade anonymity. To account for arbitrary dependencies and correlations of the errors within stocks and on particular days we could double cluster the standard errors by stock and by date as per Petersen (2009) and Thompson (2011). The variable of interest is the coefficient β, which measures the effect of post-trade anonymity on the microstructure metric y_{it}. Potential modifications to the above model include replacing the stock and time fixed effects with the dummy variables, $D_i^{AnonStock}$ and $D_t^{AnonPeriod}$, and/or adding additional control variables.

REFERNCES

Amihud, Y., 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31-56.

Anand, A., V.A. Gatchev, L. Madureira, C.A. Pirinsky, and S. Underwood, 2011, Geometric proximity and price discovery: Evidence from NASDAQ, *Journal of Financial Markets* 14, 193-226.

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

Ates, A., and G.H.K. Wang, 2005, Information transmission in electronic versus open-outcry trading systems: An analysis of US euity index futures markets, *Journal of Futures Markets* 25, 679-715.

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

Bessembinder, H., 2003, Issues in assessing trade execution costs, *Journal of Financial Markets* 6, 233-257.

Black, F., 1971, The only game in town, *Financial Analysts Journal* 27, 29-34.

Boehmer, E. and J. Wu, 2013, Short selling and the price discovery process, *Review of Financial Studies* 26, 287-322.

Booth, G.G., R.W. So, and Y. Tse, 1999, Price discovery in the German equity index derivatives markets, *Journal of Futures Markets* 19, 619–643.

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

Bourghelle, D., and F. Declerck, 2004, Why markets should not necessarily reduce the tick size, *Journal of Banking and Finance* 28, 373-398.

Cabrera, J., T. Wang, and J. Yang, 2009, Do futures lead price discovery in electronic foreign exchange markets?, *Journal of Futures Markets* 29, 137-156.

Cao, C., O., Hansch, and X. Wang, 2009, The information content of an open limit-order book, *Journal of Futures Markets* 29, 16-41.

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

Chan, K.C., W.G. Christie, and P.H. Schultz, 1995, Market structure and the intraday pattern of bid-ask spreads for NASDAQ securities, *Journal of Business* 68, 35-60.

Chen, W.-P., and H. Chung, 2012, Has the introduction of S&P 500 ETF options led to improvements in price discovery in SPDRs?, *Journal of Futures Markets* 32, 683-711.

Chen, Y.-L., and Y.-F. Gau, 2010, News announcements and price discovery in foreign exchange spot and futures markets, *Journal of Banking and Finance* 34, 1628-1636.

Chen, H., and P.M. Sub Choi, 2012, Does information vault Niagara Falls? Cross-listed trading in New York and Toronto, *Journal of Empirical Finance* 19, 175-199.

Chen, H., P.M. Sub Choi, and Y. Hong, 2013, How smooth is price discovery? Evidence from cross-listed stock trading, *Journal of International Money and Finance* 32, 668-699.

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

Chordia, T., R. Roll, and A. Subrahmanyam, 2005, Evidence on the speed of convergence to market efficiency, *Journal of Financial Economics* 76, 271-292.

Chu, Q.C., W.G. Hsieh, and Y. Tse, 1999, Price discovery on the S&P 500 index markets: An analysis of spot index, index futures and SPDRs, *International Review of Financial Analysis* 8, 21–34.

Covrig, V., D.K. Ding, and B.S. Low, 2004, The contribution of a satellite market to price discovery: Evidence from the Singapore Exchange, *Journal of Futures Markets* 24, 981-1004.

Engle, R.F., and J. Lange, 2001, Predicting VNET: A model of the dynamics of market depth, *Journal of Financial Markets* 4, 113-142.

Eun, C.S., and S. Sabherwal, 2003, Cross-border listings and price discovery: Evidence from US-listed Canadian stocks, *Journal of Finance* 58, 549-575.

Forte, S., and J.I. Pena, 2009, Credit spreads: An empirical analysis of the informational content of stocks, bonds, and CDS, *Journal of Banking and Finance* 33, 2013-2025.

Fong, K., and R. Zurbruegg, 2003, How much do locals contribute to the price discovery process? *Journal of Empirical Finance* 10, 305-320.

Fricke, C., and L. Menkhoff, 2011, Does the "Bund" dominate price discovery in Euro bond futures? Examining information shares, *Journal of Banking and Finance* 35, 1057-1072.

Frijns, B., A. Gilbert, and A. Tourani-Rad, 2010, The dynamics of price discovery for cross-listed shares: Evidence from Australia and New Zealand, *Journal of Banking and Finance* 34, 498-508.

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

Goldstein, M.A., and K.A Kavajecz, 2000, Eighths, sixteenths, and market depth: Changes in tick size and liquidity provision on the NYSE, *Journal of Financial Economics* 56, 125-149.

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

Harris, F.H.deB., T.H. McInish, and R.A. Wood, 2002, Security price adjustment across exchanges: An investigation of common factor components in Dow stocks, *Journal of Financial Markets* 5, 277-308.

Harris, M., and A. Raviv, 1993, Differences of opinion make a horse race, *Review of Financial Studies* 5, 473-506.

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

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

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

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

Hsieh, W.-L.G., C.-S. Lee, and S-F. Yuan, 2008, Price discovery in the options market: An application of put-call parity, *Journal of Futures Markets* 28, 354-375.

Huang, 2002, The quality of ECN and Nasdaq market maker quotes, *Journal of Finance* 57, 1285-1319.

Keim, D.B., and A. Madhavan, 1996, The upstairs market for large-block transactions: Analysis and measurement of price effects, *Review of Financial Studies* 9, 1-36.

Korczak, P., and K. Phylaktis, 2010, Related securities and price discovery: Evidence from NYSE-listed non-US stocks, *Journal of Empirical Finance* 17, 566-584.

Kurov, A., and D.J. Lasser, 2004, Price dynamics in the regular and E-Mini futures markets, *Journal of Financial and Quantitative Analysis* 39, 365-384.

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

Lee, C.M.C., and M.J. Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 46, 733–746.Lehmann, B.N., 2002, Some desiderata for the measurement of price discovery across markets, *Journal of Financial Markets* 5, 259–276.

Lehmann, B.N., 2002, Some desiderata for the measurement of price discovery across markets, *Journal of Financial Markets* 5, 259–276.

Lin, J.-C., G.C. Sanger, and G.G. Booth, 1995, Trade size and components of the bid-ask spread, *Review of Financial Studies* 8, 1153-1183.

Liu, Q., and Y. An, 2011, Information transmission in informationally linked markets: Evidence from US and Chinese commodity futures markets, *Journal of International Money and Finance* 30, 778-795.

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

McInish, T.H., and R.A. Wood, 1992, An analysis of intraday patterns in bid/ask spread for NYSE stocks, *Journal of Finance* 47, 753-764.

Michayluk, D. and K. Neuhauser, 2008, Is liquidity symmetric? A study of newly listed internet and technology stocks, *International Review of Finance* 8, 159-178.

Mizrach, B., and C.J. Neely, 2008, Information shares in the US Treasury market, *Journal of Banking and Finance* 32, 1221-1233.

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

Naes, R., and J.A. Skjeltorp, 2006, Order book characteristics and the volume-volatility relation: Empirical evidence from a limit order market, *Journal of Financial Markets* 9, 408-432.

O'Hara, M., 1995, *Market microstructure theory*, (Blackwell Publishers: Cambridge).

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

Pascual, R., B. Pascual-Fuster, and F. Climent, 2006, Cross-listing, price discovery and the informativeness of the trading process, *Journal of Financial Markets* 9, 144-161.

Pastor, L., and R. Stambaugh, 2003, Liquidity risk and expected returns, *Journal of Political Economy* 111, 642-685.

Pesaran, H.H., and Y. Shin, 1998, Generalized impulse response analysis in linear multivariate models, *Economics Letters* 58, 17–29.

Petersen, M.A., 2009, Estimating standard errors in finance panel data sets: Comparing approaches, *Review of Financial Studies* 22, 435–480.

Putnins, T.J., 2013, What do price discovery metrics really measure?, *Journal of Empirical Finance* 13, 3–33.

Rittler, D., 2012, Price discovery and volatility spillovers in European Union emissions trading scheme: A high-frequency analysis, *Journal of Banking and Finance* 36, 774-785.

Shastri, K., R.S. Thirumalai, and C.J. Zutter, 2008, Information revelation in the futures market: Evidence from single stock futures, *Journal of Futures Markets* 28, 335-353.

Thompson, S.B., 2011, Simple formulas for standard errors that cluster by both firm and time, *Journal of Financial Economics* 99, 1-10.

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