Empirical Market Microstructure Analysis (EMMA)

Lecture 4: Transaction and Trading Models (1/2)

Prof. Dr. Michael Stein
michael.stein@vwl.uni-freiburg.de

Albert-Ludwigs-University of Freiburg

Summer Term 2016
Outline

1. Introduction: Financial Markets and Market Structure
3. Statistical Building Blocks and Econometric Basics
4. Transaction and Trading Models
5. Information-Based Models
6. Inventory Models
7. Limit Order Book Models
8. Price Discovery and Liquidity
9. High-frequency Trading
10. Current Developments
Important Notes

- The first 3 lectures provided the groundwork for the specific models on market microstructure and the empirical analyses regarding the testing of those.

- Lecture 4 provides several definitions of the bid-ask spread that are needed when aiming at estimating models of financial market microstructure. The bid-ask spread is in the center of many empirical applications of financial market microstructure.

- After the definitions are introduced, we focus on the Roll model, one of the most popular models in market microstructure. We will derive the model and discuss its properties.

- Lecture 5 will be devoted to transactions and trading models as well, where we focus on the Choi, Salandro and Shastri extension and on the popular Glosten-Harris model which are related to the Roll model. Following the general discussion, empirical results will be discussed.
Needed Definitions

• The bid-ask spread:
  • Given the highest price a buyer is willing to pay and the lowest price a seller would accept, we obtain the bid-ask spread as the difference between those, i.e. the bid-ask spread is the difference between bid (sell price) and ask (buy price).
  • Components of the bid-ask spread generally are: adverse selection costs, fixed costs and inventory costs.
  • The bid-ask spread is the most researched measure in financial market microstructure econometrics.
Components of the bid-ask spread

- Adverse selection costs arise from asymmetric information. This corresponds to classical microeconomic reasonings of asymmetric information.
- Fixed costs, or order processing costs, stem from administrative costs or compensations for market-makers.
- Inventory costs derive from the fact that market-makers may need to hold positions other than that they would otherwise want to hold, and change bid and ask to induce offsetting trades.
- In this lecture we will focus on the order-processing component of the bid-ask spread.
Transaction Costs and the bid-ask spread

- There are several definitions of the bid-ask spread and accordingly of the transaction costs. We will follow De Jong and Rindi in their discussion of the various measures before discussing the prominent Roll model for spread estimation in detail.

- Generally, the bid-ask spread serves as an indicator of liquidity - the smaller the spread, the higher the liquidity and vice versa.

- Bid-ask spreads are a measure of trading costs.

- Implicit measure of transaction costs as opposed to explicit transaction costs like brokerage fees, stamp duties or taxes.

- Various definitions exist for the spread, like quoted spread, effective spread and realized spread.
Quoted Spread

- The quoted spread is the cost an investor would have to incur for a round-trip transaction, i.e., a purchase and sale of an asset:

\[ S^Q = \frac{1}{T} \sum_{t=1}^{T} (A_t - B_t) \]

with \( A_t \) being the ask and \( B_t \) being the bid price.

- From this, we obtain the relative quoted spread by dividing with the mid-point of bid and ask:

\[ s^Q = \frac{1}{T} \sum_{t=1}^{T} \frac{A_t - B_t}{(A_t + B_t)/2} \]
Quoted Spread

- Alternatively, the relative quoted spread may be obtained by using the natural logarithm of the bid and ask:

\[ S^Q = \frac{1}{T} \sum_{t=1}^{T} (\ln A_t - \ln B_t) \]

- The quoted spread has the disadvantage, that it may be biased to the downside when during the busiest time a small spread is prevalent (happens rather often). Many observations have small spreads during that time, while there may be larger spreads during other trading times.

- While this may be healed with weighing the average using time durations, other problems remain: For example the quoted spread may not be binding or is for very small volumes. This leads to preference of definitions that are based on actually done trades.
**Definition 2:** The *transaction costs* at time \( t \) are equal to:

\[
Q_t (P_t - P_t^*)
\]  

(6.3)

where \( P_t^* \) is equal to the equilibrium price, or fundamental value, and \( P_t \) denotes the transaction price at \( t \). The variable \( Q_t \) indicates whether the transaction was buyer-initiated (‘buy’, \( Q_t = 1 \)) or seller-initiated (‘sell’, \( Q_t = -1 \)).

In general, ‘transaction cost’ is the cost of a single transaction (buy or sell), whereas spread is defined as the costs associated with a round-trip transaction, i.e. a purchase followed by a sale of the same amount (assuming the equilibrium price does not change). From this, a natural estimator for the spread is:

\[
S = \frac{1}{T} \sum_{t=1}^{T} 2Q_t(P_t - P_t^*)
\]  

(6.4)

where \( T \) is the number of observations over a given period. Notice that taking the price \( P_t \) denominated in a particular currency yields the quoted spread in currency units, whereas taking the logarithm (\( \ln P_t \)) gives the relative spread, which is preferable when several stocks are compared.

The fundamental value of the security \( P_t^* \) is typically not observable, but the transaction prices \( P_t \) are. We must introduce a definition of spread as a function of observable variables. An operational measure of spread is the absolute difference between the transaction price and the midpoint of bid and ask quotes. This is usually called the effective spread.

**Source:** De Jong & Rindi
Effective Spread

• Given that we want to estimate the spread from actual transactions rather than quotes, we use the transaction price and compare it to the midpoint of bid and ask:

\[ S^E = \frac{1}{T} \sum_{t=1}^{T} 2Q_t(P_t - M_t), \quad M_t = \frac{A_t + B_t}{2} \]

• The variable \( Q_t \) is an indicator of the initiation direction of the trade: 1 for buy and 1 for sell.

• As for the quoted spread, the use of natural logarithms gives the relative counterpart:

\[ s^E = \frac{1}{T} \sum_{t=1}^{T} (\ln P_t - \ln M_t) \]
Effective Spread

- As data on the initiation may not be available, the absolute difference may be used for the derivation of the effective spread:

\[ S^E = \frac{1}{T} \sum_{t=1}^{T} 2|P_t - M_t| \]

- While the measure may be valid when information is symmetric, asymmetric information may lead to price adjustments due to trades. This leads to the use of the midpoint of bid and ask after the transaction, resulting in the realized spread.
Realized Spread

- The definition is analogous to that of the effective spread, only that the midpoint after a transaction is used:

\[ S^R = \frac{1}{T} \sum_{t=1}^{T} 2Q_t(P_t - M_{t+1}) \]

- This follows from Biais, Foucault and Hillion (1997):

\[ S^R = E(\Delta P_t | P_{t-1} = B_{t-1}) - E(\Delta P_t | P_{t-1} = A_{t-1}) \]

- Normally, the realized spread is smaller than the quoted spread. To estimate the former, we only need quotes, whereas for the latter we need both quotes and transaction prices.
Example with midprice reduction

Source: De Jong & Rindi, p. 95
Effective and quoted spreads

Source: De Jong & Rindi, p. 95
Estimating bid-ask spreads

- Data on quotes may be problematic as the quotes may not be binding, so the spread may vary over time and estimation of spreads is preferably done using transaction data.

- One of empirical market microstructure’s most prominent models, the model of Roll (1994), addresses this problem with a very parsimonious approach.

- Interestingly, the model shows that under the given assumptions, quoted and realized spreads are equal if adverse selection costs and inventory costs are not present.
Roll (1984) Model

- Roll proposed a method that needs nothing more than the time series of market prices.

Source: Roll (1984)

- Before we go into the model discussion, we focus on two aspects that are important in understanding the model: the role of information in the context of efficient market hypothesis (EMH) and the relation of the model with the random walk.
Martin Sewell provides a worthwhile webpage (http://www.e-m-h.org/taxonomy.html) devoted to the efficient market hypothesis.

- A taxonomy therein provides the following:
  - Weak Form Efficiency: The information set includes only the history of prices.
  - Semi-strong Form Efficiency: The information set includes all information known to all market participants (publicly available information).
  - Strong Form Efficiency: The information set includes all information known to any market participant (private information).
Relation with random walk

• An autoregression function of order \( p \) can be expressed in terms of difference equations as well:

\[
y_t - y_{t-1} = \alpha_0 + (\alpha_1 - 1) \cdot y_{t-1} + \sum_{i=2}^{p} \alpha_i \cdot y_{t-i} + x_t
\]

\[\Rightarrow \Delta y_t = \alpha_0 + (\alpha_1 - 1) \cdot y_{t-1} + \sum_{i=2}^{p} \alpha_i \cdot y_{t-i} + x_t\]

\[\Rightarrow \Delta y_t = \alpha_0 + \gamma \cdot y_{t-1} + \sum_{i=2}^{p} \alpha_i \cdot y_{t-i} + x_t\]

• In case that \( p = 1 \), \( \alpha_0 = 0 \) and \( \alpha_1 = 1 \) (\( \gamma = 0 \)) we obtain the so-called random walk:

\[
y_t - y_{t-1} = x_t \iff \Delta y_t = x_t
\]

• Adding a drift:

\[
y_t - y_{t-1} = \mu + x_t \iff \Delta y_t = \mu + x_t
\]
Relation with random walk

- The random walk hypothesis states that the daily rates of change of a stock are expected to be zero, have a zero mean. This derives from the notion that knowledge of an increase in the price would drive prices speculatively up and vice versa. In the form of a difference equation this would result as seen above. A simulation example (no drift) what would result for the level ($y$):
Relation with random walk

- For most microstructure econometric applications, where the variable of interest is a price series, the drift can be assumed to be zero. If this is the case, it is not possible to form a forecast for the process other than the recent value:
  \[ E[p_{t+1} \mid p_t, p_{t-1}, \ldots] = p_t \]

- Discrete stochastic processes like this are called martingales. For a martingale process, the following may be defined:
  \[ E[x_{t+1} \mid x_t, x_{t-1}, \ldots] = x_t \text{ and } E \mid x_t \mid < \infty \]

- This may be generalized when conditioning is done not only on the process itself, but on broader information sets. Consider the case where the process is a martingale with respect to another process \( \{z_t\} \):
  \[ E[x_{t+1} \mid z_t, z_{t-1}, \ldots] = x_t \text{ and } E \mid x_t \mid < \infty \]

- Accordingly, for sequences of information sets \( \phi_1, \phi_2, \ldots, \phi_k \), the value of the process (the conditional expectation) at time \( t \) is a martingale with respect to the sequence of all \( k \) information sets when the terminal value is a random variable \( \nu \):
  \[ x_t = [\nu \mid \phi_t] \]
Relation with random walk

- Having defined the properties of the random walk and martingales, the conditional expectation can be seen as the fundamental value, equilibrium or efficient price of an asset, for the case that the conditioning information is “all public information”.

- Assuming a random walk for the (martingale) efficient price rather than the actual transaction price then builds the bridge to the Roll model:

> If the market is informationally efficient, and trading costs are zero, the observed market price contains all relevant information.¹ A change in price will occur if and only if unanticipated information is received by market participants. There will be no serial dependence in successive price changes (aside from that generated by serial dependence in expected returns).

> When transactions are costly to effectuate, a market maker (or dealer) must be compensated; the usual compensation arrangement includes a bid-ask spread, a small region of price which brackets the underlying value of the asset. The market is still informationally efficient if the underlying value fluctuates randomly. We might think of “value” as being the center of the spread. When news arrives, both the bid and the ask prices move to different levels such that their average is the new equilibrium value. Thus, the bid-ask average fluctuates randomly in an efficient market.

Source: Roll (1984)
Trading at bid and ask

- As we know, trading is done with bid and ask, and therefore not at the “average” or any fundamental or equilibrium price.
- Roll assumes that the spread \( s \) of the market maker is constant over time, that there is no difference in the probability of buyers and sellers arriving in the market, and that there is no new information in the market.
- Given these assumptions, the bid and ask fluctuates around the value of the asset.
Trading at bid and ask

• Below we can see that the model allows for any path of trading at bid and ask, here following a sale at the market makers‘ bid:

Source: Roll (1984)
The two parts of the probability distribution show the elegant way to derive possible price changes when there is no new information and therefore a price increase (decr.) cannot follow a price increase (decr.):
Resulting price change probabilities

- As arrivals of buyers and sellers are equally probable, the combined joint distribution of price changes is:

\[
\begin{array}{|c|c|c|}
\hline
\Delta p_{t+1} & \Delta p_t & \\
\hline
-s & 0 & +s \\
0 & 1/8 & 1/8 \\
1/8 & 1/4 & 1/8 \\
s & 1/8 & 0 \\
\hline
\end{array}
\]

- Note that while there may be no price increases followed by price increases, and no price decreases followed by price decreases, it is indeed possible that buying follows buying and that selling follows selling. Bid and ask however stay the same thereby.
Formal representations

- De Jong and Rindi summarize the technical aspects of the model in very comfortable fashion:
  - Probabilities of buying and selling are equal:
    \[ Pr(Q_t = 1 | \Phi_{t-1}) = (Q_t = -1 | \Phi_{t-1}) = \frac{1}{2} \]
  - Random walk for the equilibrium price and assuming that the equilibrium price is the transaction price would yield:
    \[ \Delta P_t^* = P_t^* - P_{t-1}^* = \Gamma + U_t \]
where \( \Gamma \) is a constant equal to the unconditional expected value of \( \Delta M_t \) in the time interval \( \Delta t \), corresponding to the expected variation in the equilibrium price, and \( \Delta U_t \) is a random variable with zero mean and variance \( \sigma_U^2 \), representing the revision of the equilibrium price generated in the period \( \Delta t \) by the unexpected arrival of public information.
Formal representations

- Constant spread $S$ added to efficient or equilibrium price and dropping the drift:

$$P_t = P_t^* + \frac{S}{2}Q_t$$

- Resulting price change:

$$\Delta P_t = P_t^* - P_{t-1}^* + \frac{S}{2}(Q_t - Q_{t-1})$$

$$= U_t + \frac{S}{2}\Delta Q_t$$
Formal representations

- Analogous to the probability table for price changes in Roll (1984), De Jong and Rindi provide the probability table for the change in trade direction:

<table>
<thead>
<tr>
<th>( \Delta Q_{t-1} )</th>
<th>-2</th>
<th>0</th>
<th>+2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta Q_t = -2 )</td>
<td>0</td>
<td>\frac{1}{8}</td>
<td>\frac{1}{8}</td>
</tr>
<tr>
<td>( \Delta Q_t = 0 )</td>
<td>\frac{1}{8}</td>
<td>\frac{1}{4}</td>
<td>\frac{1}{8}</td>
</tr>
<tr>
<td>( \Delta Q_t = +2 )</td>
<td>\frac{1}{8}</td>
<td>\frac{1}{8}</td>
<td>0</td>
</tr>
</tbody>
</table>

Source: De Jong & Rindi, p. 98
Formal representations

- This stems from the following tree of possible paths of trading directions:
Formal representations

- And from the following tree of possible paths of trading direction changes:

$$\Delta Q_{t-1} \quad \quad \quad \quad \quad \quad \Delta Q_t$$

- $0$  $\rightarrow$  $0$

- $-2$  $\rightarrow$  $-2$

- $2$  $\rightarrow$  $2$

- $0$  $\rightarrow$  $0$
Joint probabilities

• We can see that there is a 50% chance that the direction of trade is different to the period before.

• However, the notion that De Jong and Rindi take, that an upturn is most likely followed by a downturn, is by construction, rather than an insight. This is because the 4 paths buy-buy, buy-sell, sell-buy and sell-sell have three different outcomes, with buy-sell and sell-buy both leading to the result of 0.

• It is easy to multiply the outcomes with the probabilities, yielding:

\[ E(\Delta Q_t \Delta Q_{t-1}) = -1 \]
Joint probabilities and covariance

- Concerning the price changes, we can multiply the outcomes with the probabilities as well, where the general result using Roll’s joint probability table is:
  \[ E(\Delta p_t \Delta p_{t-1}) = -S \times S \times \frac{1}{8} + (-S \times S \times \frac{1}{8}) = -\frac{S^2}{4} < 0 \]
- This result is the auto-covariance of the prices. In the specification of a random-walk model with uncorrelated random disturbances, De Jong and Rindi specify the result as follows:

\[
\text{Cov}(\Delta P_t, \Delta P_{t-1}) = \text{Cov}\left[\left(U_t + \frac{S}{2} \Delta Q_t\right), \left(U_{t-1} + \frac{S}{2} \Delta Q_{t-1}\right)\right] \\
= -\frac{S^2}{4} < 0
\]
Calculating the spread

- Both the general form and the one explicitly taking into account the random-walk property yield the result of a negative covariance between successive price observations. From this result, the spread may be derived as:

\[ S(Roll) = 2\sqrt{-\text{Cov}(\Delta P_t, \Delta P_{t-1})} \]

- It can be shown that the variance of the price change is \( S^2 \) and the resulting autocorrelation is \(-\frac{1}{2}\). Roll provides the intuition for this result:

> The magnitude of this autocorrelation coefficient might appear to be implausible because much smaller (in absolute value) autocorrelations are invariably found in asset returns; cf., Fama [3], the original and classic article on the subject. But observed autocorrelation coefficients may be small because the covariance is divided by the sample variance of unconditional price changes. The variance of observed price changes is likely to be dominated by new information, whereas the covariance between successive price changes cannot be due to new information if markets are efficient.\(^2\) The large new information component in the observed sample variance results in small observed serial correlation coefficients. Thus, in attempting to measure the bid-ask spread, we would be well-advised to work only with serial covariances, not with autocorrelations or with variances since these latter statistics are polluted (for present purposes) by news.

Source: Roll (1984)
Calculating the spread

• Empirical tests may be done already directly on the price change, with the formula from above yielding the regression setup with the slope coefficient being half the spread:

\[
\Delta P_t = P^*_t - P^*_{t-1} + \frac{S}{2} (Q_t - Q_{t-1})
\]

\[
= U_t + \frac{S}{2} \Delta Q_t
\]

\[
\Delta P_t = \beta_0 + \beta_1 \Delta Q_t + e_t
\]

• However, this would make the availability of direction data necessary in addition to the price data. If such data is unavailable, the Roll estimator derived above is employed.
Calculating the spread

• Roll used data of stocks listed at the NYSE and AMEX, from 1963 to 1982. From the results above, it is straightforward that the spread was measured as a function of the autocovariance:

\[ \hat{s}_{j,t} = 200 \sqrt{-\hat{c}_{j,t}} \]

• Larger firms are traded with more volume and more volume normally induces smaller spreads, comparison needed.

• Roll took care of possible misspecifications by using cross-sectional measures as well, and compared the results on daily and weekly basis.
Empirical results

Table I
Estimated Bid-Ask Spread and Size," AMEX and NYSE Listed Stocks, 1963–82, One-Day (Daily) and Five-Day (Weekly) Returns

<table>
<thead>
<tr>
<th>Year</th>
<th>Sample Size</th>
<th>Cross-Sectional Mean Spread $s$ (%), (t-statistic)</th>
<th>Cross-Sectional Regression $s_{it} = a + b \log_{10}(Size_{it-1})$, (t-statistic)</th>
<th>Cross-Sectional Rank Correlation of $s_{it}$ and $Size_{it-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One-Day</td>
<td>Five-Day</td>
<td>One-Day</td>
<td>Five-Day</td>
</tr>
<tr>
<td></td>
<td>2043</td>
<td>2008</td>
<td>1.07 (20.1)</td>
<td>2.18 (26.0)</td>
</tr>
<tr>
<td></td>
<td>2061</td>
<td>2036</td>
<td>0.829 (16.5)</td>
<td>2.06 (24.6)</td>
</tr>
<tr>
<td></td>
<td>2113</td>
<td>2085</td>
<td>0.423 (8.93)</td>
<td>1.89 (22.2)</td>
</tr>
<tr>
<td></td>
<td>2150</td>
<td>2116</td>
<td>0.193 (4.29)</td>
<td>2.08 (23.7)</td>
</tr>
<tr>
<td></td>
<td>2181</td>
<td>2146</td>
<td>−0.0726 (−1.79)</td>
<td>1.69 (18.5)</td>
</tr>
<tr>
<td></td>
<td>2173</td>
<td>2127</td>
<td>−0.605 (−18.0)</td>
<td>1.74 (19.1)</td>
</tr>
<tr>
<td></td>
<td>2194</td>
<td>2154</td>
<td>−0.468 (−14.8)</td>
<td>1.58 (19.5)</td>
</tr>
<tr>
<td></td>
<td>2297</td>
<td>2269</td>
<td>−0.343 (−7.83)</td>
<td>3.05 (28.2)</td>
</tr>
<tr>
<td></td>
<td>2406</td>
<td>2373</td>
<td>−0.123 (−3.22)</td>
<td>0.838 (9.98)</td>
</tr>
<tr>
<td></td>
<td>2540</td>
<td>2516</td>
<td>0.0553 (1.43)</td>
<td>1.60 (20.7)</td>
</tr>
<tr>
<td></td>
<td>2678</td>
<td>2642</td>
<td>0.355 (6.78)</td>
<td>1.25 (12.3)</td>
</tr>
<tr>
<td></td>
<td>2707</td>
<td>2669</td>
<td>1.22 (19.3)</td>
<td>3.32 (30.0)</td>
</tr>
<tr>
<td></td>
<td>2657</td>
<td>2624</td>
<td>0.742 (13.4)</td>
<td>2.21 (19.9)</td>
</tr>
</tbody>
</table>

Source: Roll (1984)
### Empirical results

<table>
<thead>
<tr>
<th>1976</th>
<th>2610</th>
<th>2572</th>
<th>0.688</th>
<th>1.65</th>
<th>−0.533</th>
<th>−0.471</th>
<th>−0.443</th>
<th>−0.220</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(15.6)</td>
<td>(19.0)</td>
<td>(−26.1)</td>
<td>(−10.6)</td>
<td>(−25.2)</td>
<td>(−11.4)</td>
</tr>
<tr>
<td>1977</td>
<td>2570</td>
<td>2542</td>
<td>0.834</td>
<td>2.03</td>
<td>−0.565</td>
<td>−0.518</td>
<td>−0.498</td>
<td>−0.290</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(20.2)</td>
<td>(27.8)</td>
<td>(−30.0)</td>
<td>(−13.9)</td>
<td>(−29.1)</td>
<td>(−15.3)</td>
</tr>
<tr>
<td>1978</td>
<td>2515</td>
<td>2468</td>
<td>0.221</td>
<td>0.729</td>
<td>−0.368</td>
<td>−0.410</td>
<td>−0.255</td>
<td>−0.160</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(5.23)</td>
<td>(8.42)</td>
<td>(−16.9)</td>
<td>(−8.83)</td>
<td>(−13.2)</td>
<td>(−8.07)</td>
</tr>
<tr>
<td>1979</td>
<td>2439</td>
<td>2394</td>
<td>0.315</td>
<td>1.00</td>
<td>−0.417</td>
<td>−0.483</td>
<td>−0.311</td>
<td>−0.214</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(7.69)</td>
<td>(12.1)</td>
<td>(−19.7)</td>
<td>(−10.8)</td>
<td>(−16.1)</td>
<td>(−10.7)</td>
</tr>
<tr>
<td>1980</td>
<td>2375</td>
<td>2328</td>
<td>0.265</td>
<td>1.11</td>
<td>−0.551</td>
<td>−0.250</td>
<td>−0.390</td>
<td>−0.0939</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(5.62)</td>
<td>(12.5)</td>
<td>(−23.4)</td>
<td>(−5.10)</td>
<td>(−20.6)</td>
<td>(−4.55)</td>
</tr>
<tr>
<td>1981</td>
<td>2348</td>
<td>2294</td>
<td>0.170</td>
<td>1.38</td>
<td>−0.504</td>
<td>−0.238</td>
<td>−0.358</td>
<td>−0.111</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2.90)</td>
<td>(16.7)</td>
<td>(−16.8)</td>
<td>(−5.34)</td>
<td>(−18.5)</td>
<td>(−5.34)</td>
</tr>
<tr>
<td>1982</td>
<td>2339</td>
<td>2277</td>
<td>−0.0492</td>
<td>1.41</td>
<td>−0.418</td>
<td>−0.141</td>
<td>−0.309</td>
<td>−0.0565</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(−1.07)</td>
<td>(14.8)</td>
<td>(−17.4)</td>
<td>(−2.67)</td>
<td>(−15.7)</td>
<td>(−2.70)</td>
</tr>
<tr>
<td>All</td>
<td>47414</td>
<td>46658</td>
<td>0.298</td>
<td>1.74</td>
<td>−0.442</td>
<td>−0.495</td>
<td>−0.310</td>
<td>−0.226</td>
</tr>
<tr>
<td>Years</td>
<td>(24358)</td>
<td>(16225)</td>
<td>(27.9)</td>
<td>(84.3)</td>
<td>(−8.60)</td>
<td>(−12.6)</td>
<td>(−7.93)</td>
<td>(−10.7)</td>
</tr>
</tbody>
</table>

* The variables used are \( s_{jt} = 200 \sqrt{-\hat{c}_{jt}} \), the estimated bid-ask spread, where \( \hat{c}_{jt} \) is the serial-covariance of returns in year \( t \) on stock \( j \). (The sign of the covariance was preserved after taking the square root.) \( Size_{t-1} \) is the market capitalization in $ millions (number of shares times the price) on the last trading day of year \( t - 1 \). Stocks were discarded from the sample in year \( t \) unless they had at least 21 observations (21 trading days or approximately one month of data for daily returns and 21 weeks for five-day returns).

* Number of negative spread estimates.

* Means and \( t \)-statistics based on 20 time series observations of cross-sectional coefficients.

Source: Roll (1984)
Empirical Results

- Expected negative correlation of Roll estimated spread and size found
- Differences between weekly and daily results
- Broader discussion of all results and implications follows......
References

F. de Jong and B. Rindi
*The Microstructure of Financial Markets*. Chapter 6.1 and 6.2

The references in the mandatory chapter contain many interesting articles for further reading! Most recommended are Roll (1984) and Choi, Salandro and Shastri (1988)