Possible outcomes

- Separating equilibrium: if large size is large enough, informed trader will always trade large quantity
  - small trades only by U, hence no bid-ask spread for small size!
- Pooling equilibrium: I randomizes between small and large trades: hides some of his information to improve prices for large trades
  - spread for small size smaller than spread for large size

Important assumptions

- trading is anonymous
- informed traders act competitively: exploit information immediately

# Empirical Market Microstructure

## Based on de Jong and Rindi, Chapter 6

Frank de Jong

Tilburg University

## Estimation of the bid-ask spread

Assume a sample of $i = 1, .., N$ bid and ask quotes is available

- for example from intra-day stock exchange data

Average quoted bid-ask spread

$$\hat{S}_{quoted} = \frac{1}{N} \sum_{i=1}^{N} (ask_i - bid_i)$$

Drawbacks of this measure

- quotes may not be binding, or valid for small size only
- spread may vary over the day (U-shape typically)
- quoted spread may not be a good measure of actual trading costs if trading is busiest when spreads are small

# Transaction costs

A general definition of the transaction costs for a trade at time $t$ is

$$S_t = 2Q_t(P_t - P_t^*)$$

where $P_t$ is the transaction price, $P_t^*$ the equilibrium price true, and $Q_t$ the trade sign, defined by

$$Q_t = \begin{cases} +1 & \text{if trade is buy} \\ -1 & \text{if trade is sell} \end{cases}$$

Problem is that $P_t^*$ and often also $Q_t$ are not directly observable

## Effective spread (1)

Assume sample of $t = 1, .., T$ transactions is available

- transaction price $P_t$
- bid and ask quotes prevailing at the time of the trade
- quote midpoint is $M_t = \frac{1}{2}(ask_t + bid_t)$

Effective spread: replace $P^*$ by $M$

$$\hat{S}_{effective} = \frac{2}{T} \sum_{t=1}^{T} Q_t(P_t - M_t)$$

Note: sometimes it is difficult to measure $M_t$ at exactly the same time as the trade, creating a bias in the spread estimator

# Effective spread (2)

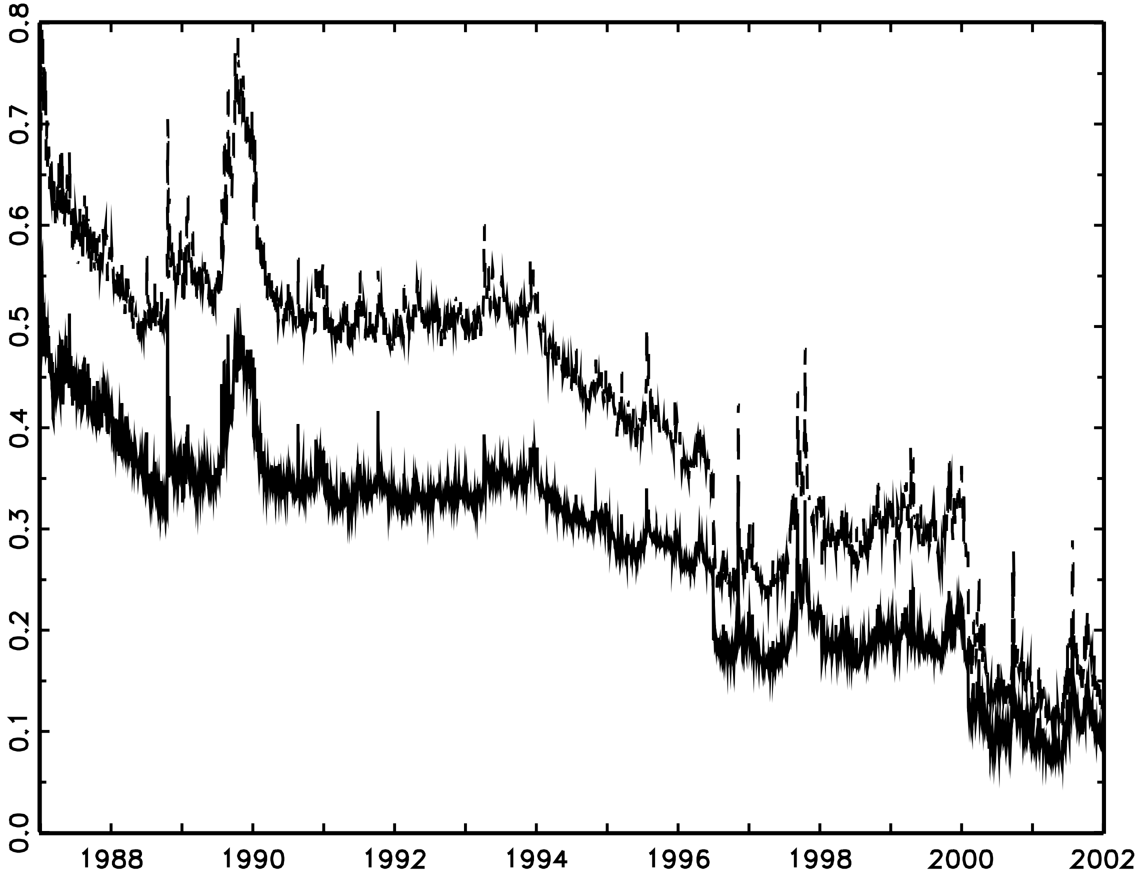
If $Q_t$ is not observed, one can assume that

- $Q_t = 1$ whenever $P_t > M_t$
- $Q_t = -1$ whenever $P_t < M_t$
- $Q_t = 0$ when $P_t = M_t$

Effective spread then is

$$\hat{S}_{effective} = \frac{2}{T} \sum_{t=1}^{T} |P_t - M_t|$$

# Quoted and Effective spread estimates




Source: Chordia, Roll and Subramanyam (2001)

# Realized spread

Trading typically has a price impact

- effective spread will overestimate cost of round-trip trade

Realized spread may be better measure

- approximate equilibrium price $P^*$ by midquote *after* the trade

$$\hat{S}_{realized} = \frac{2}{T} \sum_{t=1}^{T} Q_t (P_t - M_{t+1})$$

# Roll's estimator

Often, obtaining good quality bid and ask prices is difficult, but transaction prices are available

Roll (1977) proposed spread estimator entirely based on transaction prices

Basic idea: transaction price equals efficient price $P^*$ plus (for a buy) or minus (for a sell) half the spread

$$P_t = P_t^* + (S/2)Q_t$$

and

$$Q_t = \begin{cases} +1 & \text{if trade is buy} \\ -1 & \text{if trade is sell} \end{cases}$$

In many datasets, trade direction $Q_t$ is not observed

Assuming that $Q_t$ and $Q_{t-1}$ are uncorrelated

Define price change $\Delta P_t = P_t - P_{t-1}$; Roll showed that

$$\mathrm{Cov}(\Delta P_t,\, \Delta p_{t-1}) = -(S/2)^2$$

Re-arranging this equality gives **Roll's estimator** of the effective bid-ask spread

$$\hat{S} = 2\sqrt{-\mathrm{Cov}(\Delta P_t,\, \Delta P_{t-1})}$$

Notice that for this estimator to exist, the covariance of price changes must be negative!

- Hasbrouck (2006) suggests a Bayesian Gibbs sampling estimator top deal with this problem

Implicit assumption in Roll's estimator: trading does not affect mid-point of bid and ask quotes

# Basic microstructure model

Consider the following model of transaction prices

$$
\begin{aligned}
P_t &= P_t^* + (S/2)Q_t \\
P_{t+1}^* &= P_t^* + e_{t+1}
\end{aligned}
$$

- $P_t$ = transaction price
- $P_t^*$ = efficient price
- $Q_t$ = trade direction indicator
- $e_t$ = new public information arriving between trades
- $S$ = average bid-ask spread

# Incorporating price effects of trading

Basic microstructure model with price impact of trading

$$\begin{aligned} P_t &= P_t^* + (S/2)Q_t \\ P_{t+1}^* &= P_t^* + ZQ_t + e_{t+1} \end{aligned}$$

The new element $Z$ is the price impact of a trade

- due to adverse selection or inventory effects

Spread now consists of fixed-cost part and price impact part

$$S/2 = C + Z$$

Model can be written as

$$\begin{aligned} P_t &= P_t^* + (C + Z)Q_t \\ P_{t+1}^* &= P_t^* + ZQ_t + e_{t+1} \end{aligned}$$

Taking first differences and substituting out $P_t^*$ gives the reduced form

$$\Delta P_t = (C + Z)Q_t - CQ_{t-1} + e_t$$

The parameters $C + Z$ and $C$ can be estimated from the regression coefficients in a regression of $\Delta P_t$ on $Q_t$ and $Q_{t-1}$

An alternative way to write the regression is

$$\Delta P_t = C\Delta Q_t + ZQ_t + e_t$$

**level** variable $Q_t$ estimates adverse selection component $Z$: permanent price effect

**difference** variable $\Delta Q_t$ estimates the fixed cost component $C$: temporary price effect

# Glosten and Harris (1988)

Glosten and Harris make both spread components linear in trade size $|q|$

- temporary spread component $C = C_0 + C_1 q_t$
- permanent spread component (price impact) $Z = Z_0 + Z_1 q_t$

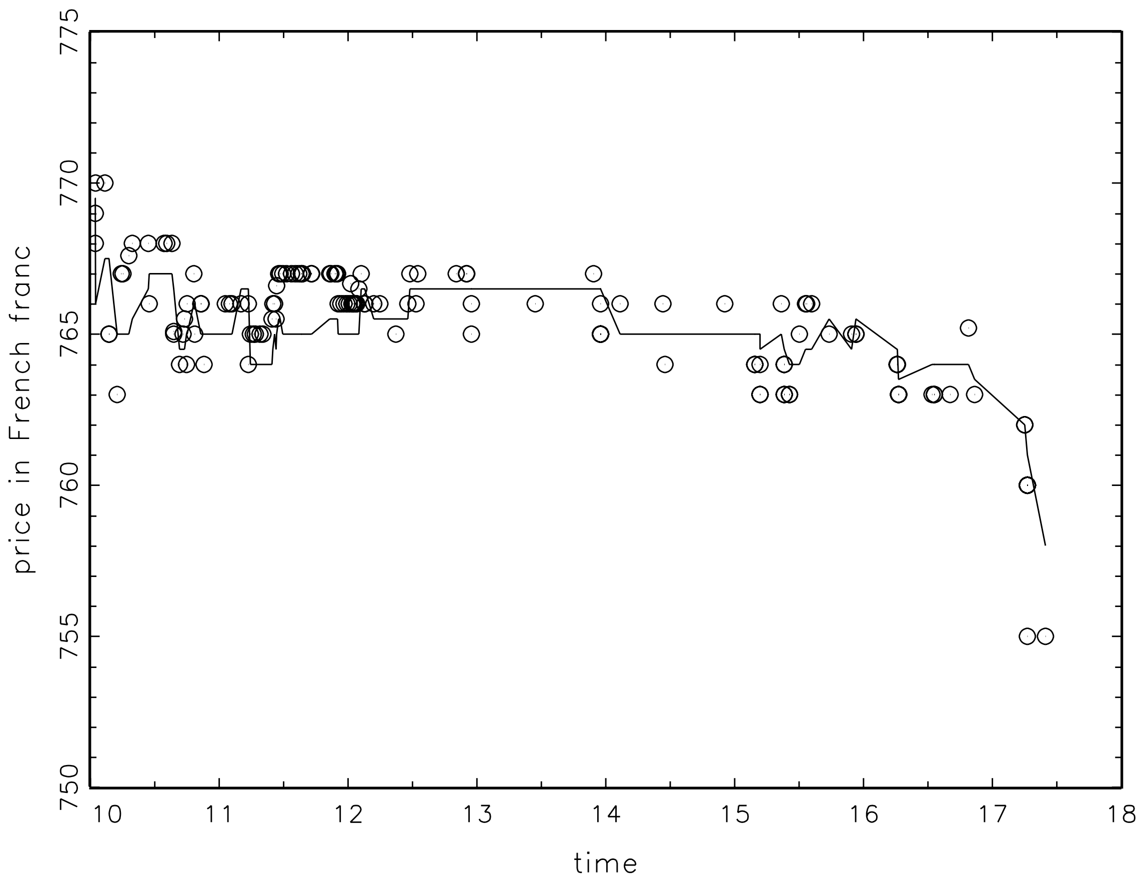
Reduced form

$$\Delta P_t = C_0 \Delta Q_t + C_1 \Delta x_t + Z_0 Q_t + Z_1 x_t + e_t$$

with $x_t = q_t Q_t$ the "signed" trade size

Data for Accor (a French hotel firm) on May 24, 1991

Transaction prices (circles) and midquote (line)



Liquidity estimates for Accor on May 24, 1991

| Measure | $S^{eff}$ | $S^{Roll}$ | $C_0$ | $Z_0$ | $C_1$ | $Z_1$ |
|---|---|---|---|---|---|---|
| Absolute (FF) | 1.02 | 0.84 | 0.71 | 0.25 | -0.07 | 0.20 |
| t-statistic | | | (16.95) | (4.88) | (-0.78) | (1.58) |
| Relative (bp) | 13.44 | 11.21 | 9.37 | 3.34 | -0.91 | 2.57 |
| t-statistic | | | (16.77) | (4.85) | (-0.77) | (1.55) |

# Amihud's $ILLIQ$ measure

- Finding data on bid-ask spreads for many stocks over long sample periods is almost impossible
- Data on returns and trading volume are readily available (e.g. from Datastream)
- Amihud suggests the following measure daily of liquidity: the absolute price change divided by trading volume for stock $i$ on day $d$

$$\frac{|R_{id}|}{V_{id}}$$

- Idea: this proxies Kyle's lambda (price impact of trading volume)

# Amihud's $ILLIQ$ measure (2)

- On any day, this ratio is a very noisy measure and in practice it is averaged over all trading days in a month or year to get a monthly or annual liquidity estimate for stock $i$

$$ILLIQ_{it} = \frac{1}{D_t} \sum_{d=1}^{D_t} \frac{|R_{id}|}{V_{id}}$$

where $D_t$ is the number of trading days in month or year $t$

- Amihud (2002) shows that $ILLIQ$ is strongly correlated with more precise measures of liquidity such as bid-ask spreads and price impacts (estimated on transactions data by the Glosten-Harris method)

- $ILLIQ$ is therefore a useful measure of illiquidity and trading costs

# Spearman rank correlations between liquidity proxies

| | Effective spread | Roll's estimator | Gibbs estimator | ILLIQ |
|---|---|---|---|---|
| Effective spread | 1.000 | | | |
| Roll's estimator | 0.636 | 1.000 | | |
| Gibbs estimator | 0.872 | 0.791 | 1.000 | |
| ILLIQ | 0.937 | 0.592 | 0.778 | 1.000 |

Source: Hasbrouck (2006), Table 2.

# The PIN model

$$
\begin{aligned}
P(news) &= \alpha \begin{cases} P(good\ news|news) = \delta \\ P(bad\ news|news) = 1-\delta \end{cases} \\
P(no\ news) &= 1-\alpha
\end{aligned}
$$

Trades arrive according to Poisson processes.

The arrival rate of uninformed buy orders is $\epsilon_B$ and that of uninformed sell orders $\epsilon_S$.

On days with information, informed trades arrive at rate $\mu$.

# The PIN model: results

Bid-ask spread follows from Glosten-Milgrom model

$$S = \frac{\alpha\mu}{\alpha\mu + \epsilon_B + \epsilon_S} [V_H - V_L]$$

Probability of informed trade

$$PIN = \frac{\alpha\mu}{\alpha\mu + \epsilon_B + \epsilon_S}$$

and covariance between number of buy and sell orders is

$$Cov(N_B, N_S) = -\alpha^2\mu^2\delta(1-\delta)$$

# Maximum Likelihood estimation of PIN

Probability density of Poisson variable $N$ is

$$P(N = n) = e^{-\lambda} \frac{\lambda^n}{n!}$$

In PIN model, arrival rates of the number of buy ($N_B$) and sell ($N_S$) trades are

$$\begin{aligned}
E(N_B | good\ news) &= \epsilon_B + \mu, \quad E(N_S | good\ news) = \epsilon_S \\
E(N_B | bad\ news) &= \epsilon_B, \quad E(N_S | bad\ news) = \epsilon_S + \mu \\
E(N_B | no\ news) &= \epsilon_B, \quad E(N_S) = \epsilon_S
\end{aligned}$$

Based on this, the likelihood function can be constructed:

$$
\begin{aligned}
L(B,S;\psi) &= \alpha\delta e^{-(\epsilon_B+\mu)}\frac{(\epsilon_B+\mu)^B}{B!}e^{-\epsilon_S}\frac{(\epsilon_S)^S}{S!} \\
&+ \alpha(1-\delta)e^{-\epsilon_B}\frac{(\epsilon_B)^B}{B!}e^{-(\epsilon_S+\mu)}\frac{(\epsilon_S+\mu)^S}{S!} \\
&+ (1-\alpha)e^{-\epsilon_B}\frac{(\epsilon_B)^B}{B!}e^{-\epsilon_S}\frac{(\epsilon_S)^S}{S!}
\end{aligned}
$$

This likelihood can be multiplied over several days (typically, all the days of a month) with observations $(B_t, S_t)$ to obtain the likelihood function

$$
\mathbf{L}(\psi) = \prod_{t=1}^{T} L(B_t, S_t; \psi)
$$