Trading activity and liquidity supply in a pure limit order book market

An empirical analysis using a multivariate count data model

Andréas Heinen,

(joint work with **Erick Rengifo** and **Joachim Grammig**)

Department of Statistics and Econometrics

Universidad Carlos III

Objectives of the paper

- Detailed study of the trading process on a pure electronic limit order book market.
- In particular: what determines the choice of orders of different degrees of aggressiveness?

What is new about this paper?

Existing literature:

- Ordered Probit model: no time dimension
- Duration models: no multivariate dimension

Multivariate counts:

- has both dimensions
- flexible
- allows forecasting of what orders will be submitted in the next interval

Outline of presentation

- Multivariate count models
- Order-driven markets
- The Xetra trading system
- Some results
- Conclusions

Motivation

- Microstructure, high frequency data
- Duration models (ACD): univariate case
- Multivariate point processes: bivariate ACD model of Lunde and Engle (99): difficult, hard to extend to more than two series
- Need flexible patterns of cross-, auto- and contemporaneous correlation

Conditional mean: a VAR in counts . . .

- $N_{i,t}|\mathcal{F}_{t-1} \sim P(\mu_{i,t})$, $\forall i = 1, \ldots, K$,
-

$$E[N_t|\mathcal{F}_{t-1}] = \mu_t = \omega + \sum_{j=1}^{p} A_j N_{t-j} + \sum_{j=1}^{q} B_j \mu_{t-j}$$

- Various assumptions on A and B: full rank (unrestricted), a priori restrictions, reduced rank (diagonal), reduced rank + own effect

...Conditional mean: special cases

- Reduced rank and own effect: for every series the conditional mean depends on one lag of itself, one lag of the count and r factors of the cross-section of lagged counts (γ and δ are $(K \times r)$):

$$\mu_t = \omega + (diag(\alpha_i) + \gamma\delta')N_{t-1} + diag(\beta_i)\mu_{t-1}$$

- Factor model:

$$\mu_t = \omega + \gamma\delta' N_{t-1} + diag(\beta_i)\mu_{t-1} \ ,$$

Conditional distribution

- Poisson: equidispersed
- double Poisson: over-, equi- or underdispersed
- dispersion coefficient ϕ,

$$V[N_t|\mathcal{F}_{t-1}] = \sigma_t^2 = \frac{\mu_t}{\phi}$$

- density:

$$f(y,\mu,\phi) = c(\mu,\phi)\left(\phi^{\frac{1}{2}}e^{-\phi\mu}\right)\left(\frac{e^{-y}y^y}{y!}\right)\left(\frac{e^{\mu}}{y}\right)^{\phi y}$$

Copulas: general idea

- Let $H(y_1, \ldots, y_K)$ be a continuous K-variate cumulative distribution function with univariate margins $F_i(y_i)$, $i = 1, \ldots, K$, where $F_i(y_i) = H(\infty, \ldots, y_i, \ldots, \infty)$.
- There exists a function C, called copula, mapping $[0, 1]^K$ into $[0, 1]$, such that:

$$H(y_1, \ldots, y_K) = C(F_1(y_1), \ldots, F_K(y_K)) \ ,$$

-

$$\frac{\partial H(y_1, \ldots, y_K)}{\partial y_1 \ldots \partial y_K} = \prod_{i=1}^{K} f_i(y_i) \frac{\partial C(F_1(y_1), \ldots, F_K(y_K))}{\partial F_1(y_1) \ldots \partial F_K(y_K)} \ .$$

Multivariate Normal Copula

- $C(z_1, \ldots, z_K; \Sigma) = \Phi^K(\Phi^{-1}(z_1), \ldots, \Phi^{-1}(z_K); \Sigma)$
- Φ^K is the K-dimensional standard normal multivariate distribution function, Φ^{-1} is the inverse of the standard univariate normal distribution function
- $$c(z_1, \ldots, z_K; \Sigma) = \mid \Sigma \mid^{-1/2} \exp\left(\frac{1}{2}(q'(I_K - \Sigma^{-1})q\right) ,$$
- $z_i = F_i(y_i)$
- $q = (q_1, \ldots, q_K)'$ with normal scores $q_i = \Phi^{-1}(z_i)$, $i = 1, \ldots, K$.

Copulas and discrete distributions

- Use of copula based on the fact that $z_i = F_i(y_i) \sim U[0,1]$ (PITT of Fisher (1932))
- Problem: unique solution only for continuous distributions!
- Solution: use continuous extension (add a continuous random variable with support [0,1])

$$Y^* = Y + (U - 1) \,.$$

-

$$Z^* = F^*(Y^*) = F\left([Y^*]\right) + f_{[Y^*]+1} U = F(Y-1) + f_y U$$

Copulas and time series

- $L_{X_1 X_2}(\theta) = L_{X_1}(\theta_1, \theta_0) + L_{X_2}(\theta_2, \theta_0) + L_C(\theta_1, \theta_2, \theta_0, \theta_c)$
- two-step procedure of Patton (02):

$$\hat{\varphi} = \underset{\varphi \in \Phi}{\text{argmax}} \quad L_{X_1}(\theta_1, \theta_0) + L_{X_2}(\theta_2, \theta_0)$$

where $\hat{\varphi} = [\hat{\theta_0}', \hat{\theta_1}', \hat{\theta_2}']'$ and in the second stage,

$$\hat{\theta}_c = \underset{\theta_c \in \Theta_C}{\text{argmax}} \quad L_C(\theta_1, \theta_2, \theta_0, \theta_c)$$

The model . . .

- Joint density of counts $N_{i,t}$:

$$h(N_{1,t},\ldots,N_{K,t},\theta,\Sigma)=\prod_{i=1}^{K} f_{DP}(N_{i,t},\mu_{i,t},\phi_i)\cdot c(q_t;\Sigma)\ ,$$

- $f_{DP}(N_{i,t},\mu_{i,t},\phi_i)$: Double Poisson density
- c: copula density of multivariate normal,
- $\theta=(\omega,vec(A),vec(B))$.
- $q_t=(\Phi^{-1}(z_{1,t}),\ldots,\Phi^{-1}(z_{K,t}))'$

...The model

- $z_{i,t} = F^*(N^*_{i,t}) = F(N_{i,t} - 1) + f(N_{i,t}) * U_{i,t}$
- $N^*_{i,t}$: continued extension of the counts $N_{i,t}$:

$$N^*_{i,t} = N_{i,t} + (U_{i,t} - 1)\ .$$

- Finally the $U_{i,t}$ are uniform random variable, on $[0,1]$.
- Taking logs, one gets:

$$log(h_t) = \sum_{i=1}^{K} log(f_{DP}(N_{i,t}, \mu_{i,t}, \phi_i)) + log(c(q_t; \Sigma))]$$

Evaluation of the models

- Log likelihood and significance of parameters

- Standardized residuals:

  1. Equidispersed?

  2. Autocorrelated?

  3. Cross-correlated?

Order-driven markets: types of orders

- market orders: immediate execution against the book (consumes liquidity)

- limit orders: executed provided price reaches a pre-specified level (supply liquidity)

- cancelations, order revisions

- limit orders are executed according to price and time priority

Buy Market Order (BMO) for 1000 shares

Initial state of the book

Shares	Limit Buy	Limit Sell	Shares
1000	98	100	4000
2000	96	102	1000
		104	2000

Buy Market Order (BMO) for 1000 shares

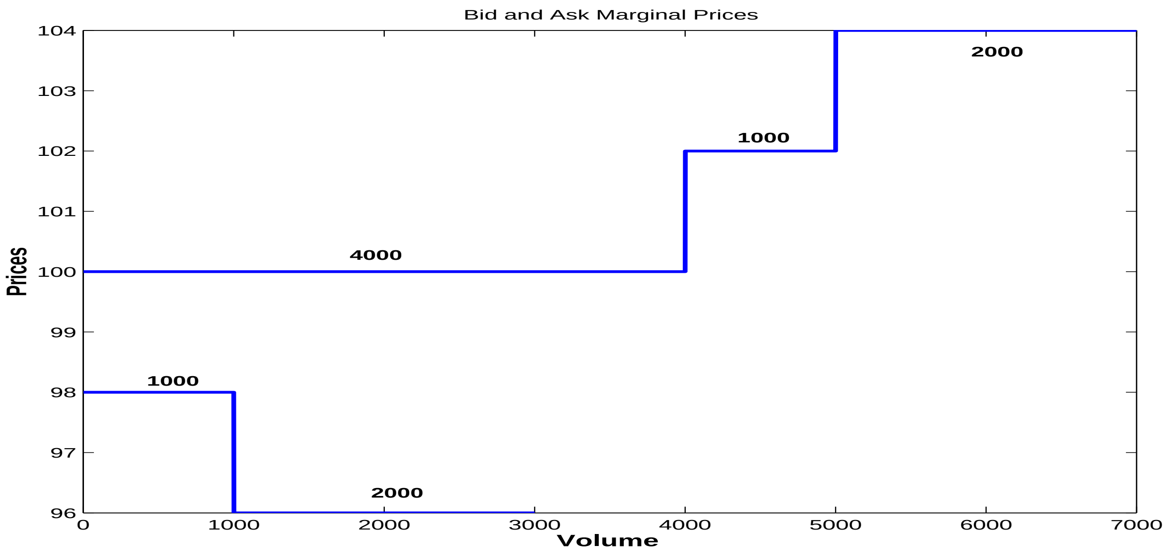
Initial state of the book

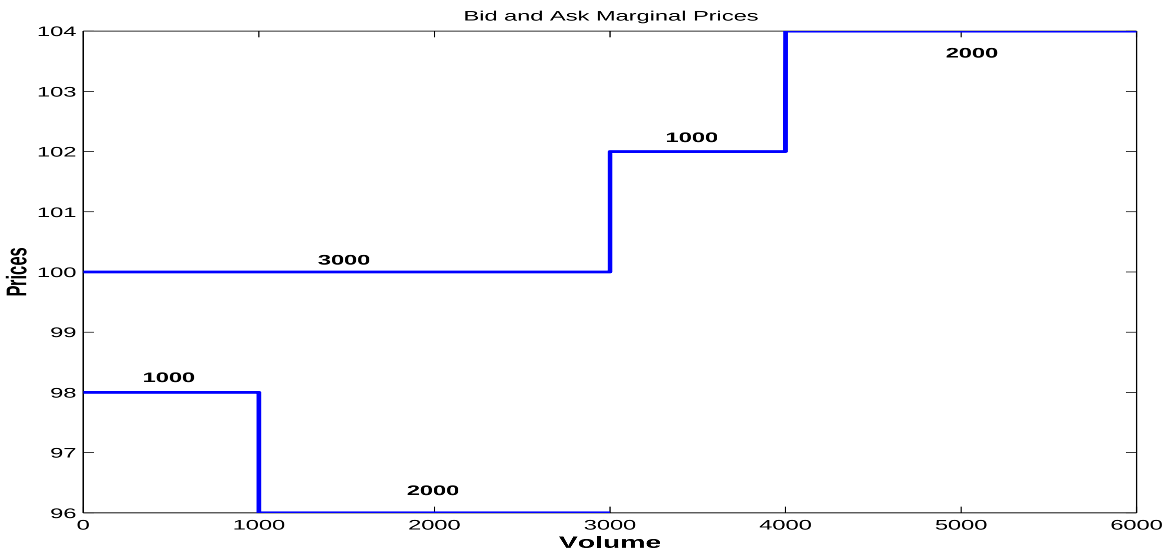
Shares	Limit Buy	Limit Sell	Shares
1000	98	100	4000
2000	96	102	1000
		104	2000

Final state of the book

Shares	Limit Buy	Limit Sell	Shares
1000	98	100	3000
2000	96	102	1000
		104	2000

Buy Market Order (BMO) for 1000 shares





Buy Limit Order (BLO) for 3000 shares, 99 Euros

Initial state of the book

Shares	Limit Buy	Limit Sell	Shares
1000	98	100	3000
2000	96	102	1000
		104	2000

Buy Limit Order (BLO) for 3000 shares, 99 Euros

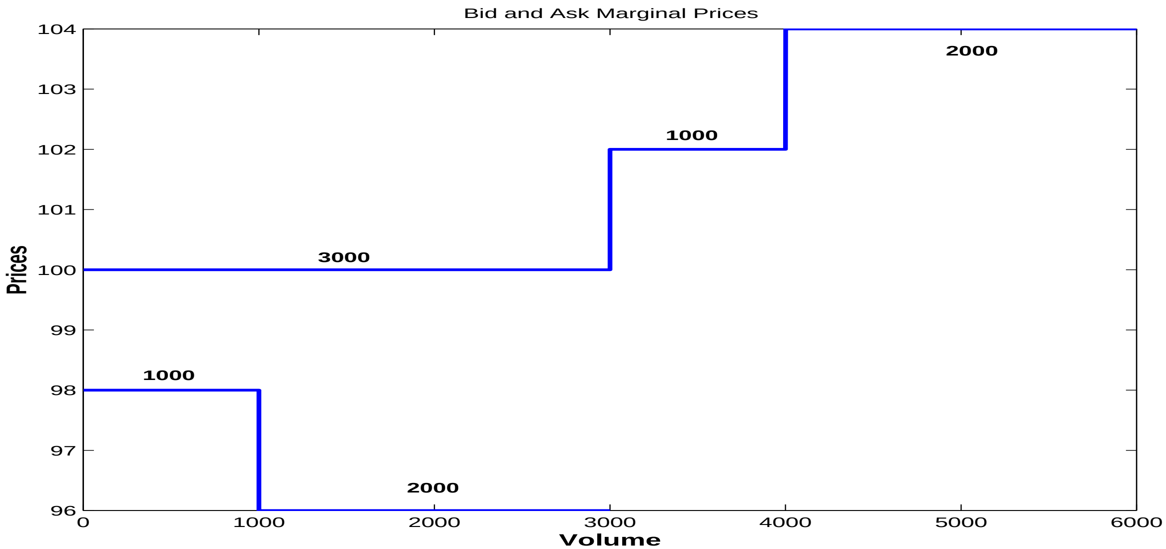
Initial state of the book

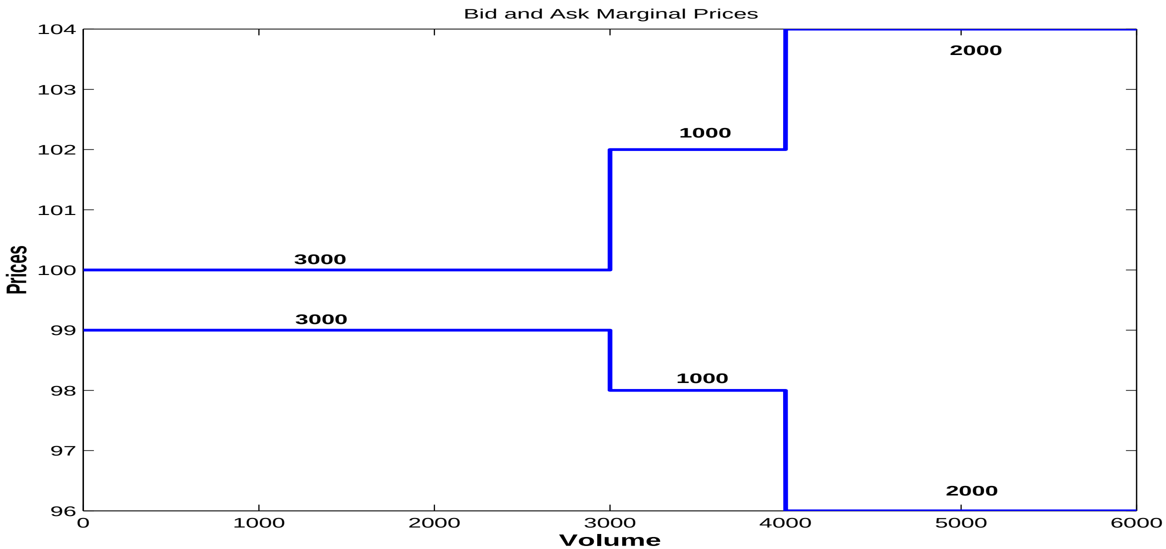
Shares	Limit Buy	Limit Sell	Shares
1000	98	100	3000
2000	96	102	1000
		104	2000

Final state of the book

Shares	Limit Buy	Limit Sell	Shares
3000	99	100	3000
1000	98	102	1000
2000	96	104	2000

Buy Limit Order (BLO) for 3000 shares, 99 Euros





Order Aggressiveness 1

- Category 1: Large market orders,
- Category 2: Market orders,
- Category 3: Small market orders,
- Category 4: Aggressive limit orders,
- Category 5: Limit orders submitted at the best quote.
- Category 6: Limit orders outside the best quotes,
- Category 7: Cancellations.

Order Aggressiveness 2

- Type 1: Between first two steps,
- Type 2: Between third and fifth steps,
- Type 3: The rest.

Buy Market Order for 5000 shares (walks up)

Initial state of the book

Shares	Limit Buy	Limit Sell	Shares
3000	99	100	3000
1000	98	102	1000
2000	96	104	2000

Buy Market Order for 5000 shares (walks up)

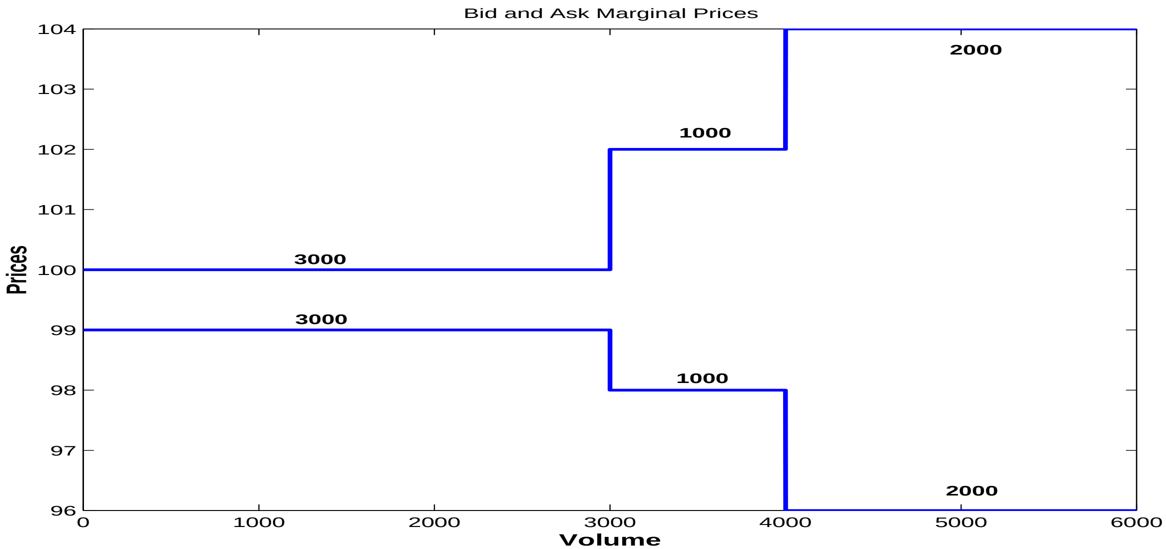
Initial state of the book

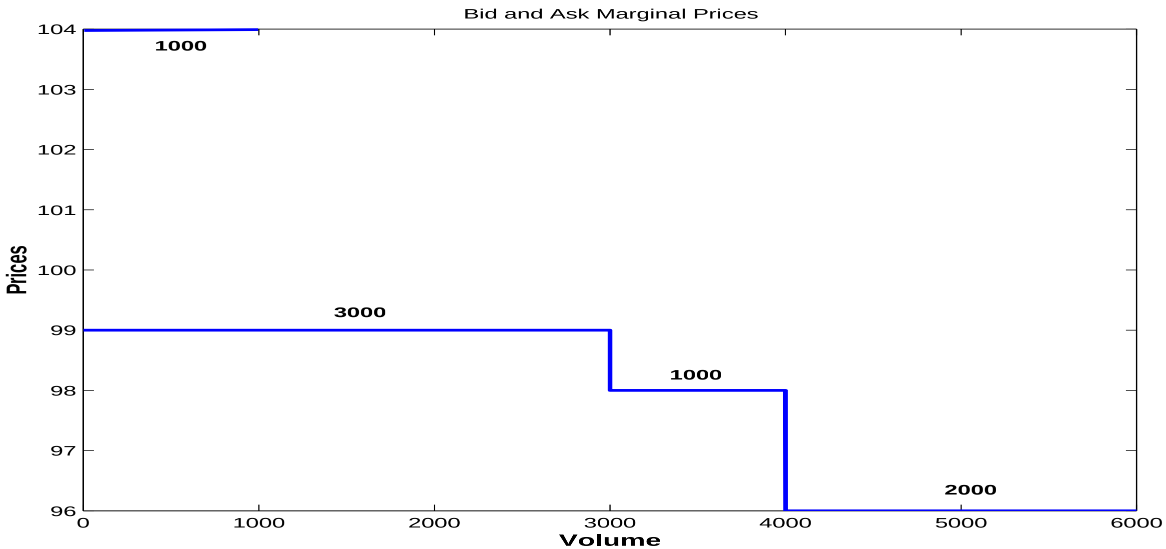
Shares	Limit Buy	Limit Sell	Shares
3000	99	100	3000
1000	98	102	1000
2000	96	104	2000

Final state of the book

Shares	Limit Buy	Limit Sell	Shares
3000	99	104	1000
1000	98		
2000	96		

Buy Market Order for 5000 shares (walks up)





The Xetra trading system

- Main trading platform of German blue chips.
- No market makers (pure limit order market).
- Continuous double auction mechanism.
- Transparent (Whole Limit Order book is displayed).
- No hidden orders (Icebergs).
- Anonymous and dual capacity trading.
- Orders can walk-up the book (MLO).
- Tick size: 0.01 Euro cents.

Data

- Assets (30.4% market-cap of DAX index):
  - Daimler Chrysler (DCX),
  - Deutsche Telekom (DTE),
  - SAP.
- Period: 20/08/99 to 20/09/99 $\longrightarrow$ 10730 minutes
- Hours: 8:45 am to 4:45 pm (only continuous trading)

Descriptive statistics of the types of orders

	Mean	Std	Disp	Max	$Q(10)$
BUY ORDERS	**4.18**	**3.55**	**3.02**	**59**	**40107**
Large MO (and MLO)	0.92	1.32	1.90	13	11177
MO	0.09	0.36	1.50	18	847
Small MO	0.24	0.54	1.22	6	8990
Total MO	**1.24**	**1.60**	**2.05**	**27**	**22403**
LO above the best bid (overbidding)	0.87	1.14	1.51	10	6246
LO at the best bid	1.06	1.33	1.68	18	14313
LO below the best bid	1.01	1.28	1.62	11	8657
Total LO	**2.94**	**2.34**	**2.55**	**32**	**23719**
Cancelation	1.91	2.03	2.15	18	13623
SELL ORDERS	**3.53**	**3.27**	**3.03**	**38**	**22260**
Large MO (and MLO)	0.68	2.62	1.95	14	8532
MO	0.06	6.22	1.20	7	568
Small MO	0.08	4.37	1.14	5	362
Total MO	**1.43**	**2.37**	**3.91**	**30**	**8890**
LO below the best ask (undercutting)	0.80	1.83	1.48	10	2818
LO at the best ask	0.95	3.06	1.78	23	8660
LO above the best ask	0.97	1.87	1.62	11	6738
Total LO	**2.71**	**2.63**	**2.54**	**38**	**15183**
Cancelations	1.86	2.09	2.34	29	11379

Determinants of order aggressiveness

- Past order flow (autoregression)
- state of the book:
  - inside spread
  - depth at the best quotes
  - factors based on the whole book
- volatility

Factors of the order book

- Deseasonalized percentage average price that an MO for volume v would pay, for various levels of v

	Factor 1	Factor 2	Factor 3
Eigenvalue	32.83	3.90	0.80
Variance Share	0.864	0.103	0.021
Cumulative Share	0.864	0.967	0.988

- use factor 1
- difference of factor 2 between buy and sell side: information

Inside spread

Hypothesis

- Handa-Schwartz (96), Foucault (99) $\longrightarrow$ higher spread, more LO, less MO

Findings:

- Aggregate system: Overall slowing down, proportion MO $\downarrow$ and LO $\uparrow$.
- Disaggregate system: negative (significant) effect on MO and negative ($\sim$significant) effect on LO.

Depth at the best quote

Hypothesis

- Depth at best quote related to execution probability of LO (Parlour (98), Handa-Schwartz-Tiwari (03))

Findings:

- On same side, more aggressiveness, on opposite side, less aggressiveness (find this with disaggregated categories)

Factors based on the whole book

Hypothesis

- liquidity (Foucault (99), Handa-Schwartz-Tiwari (03)) or information?

Findings:

- factor 1: linear relationship with markup at all volumes, liquidity factor, increases aggressiveness
- factor 2: slope factor
- difference between slopes of buy and sell side: information factor

Dynamic effects . . .

- Diagonal effect (Biais, Hillion and Spatt (95))
- buys and sells move together
- buy (sell) MO: + impact on sell (buy) LO: continuous provision of liquidity
- after large MO, LO inside the best quotes to refill the book

. . . Dynamic effects

- Cancellations carry information
- positive (negative) impact on LO (MO)
- which cancellations?: closest to the quotes
- cancellations reduce aggressiveness (less MO, less undercutting/overbidding)

Conclusions

- Model is flexible and easy to estimate
- Can accommodate discreteness, overdispersion, auto- and cross-correlation
- Accommodates positive and negative contemporaneous correlation
- feasible alternative to multivariate duration models
- many theoretical predictions from microstructure models are confirmed empirically