Censored Exploration in Dark Pools

Michael Kearns
Computer and Information Science
University of Pennsylvania

Joint work with Kuzman Ganchev, Yuriy Nevmyvaka and Jenn Wortman Vaughan

Schonfeld Group Quantitative Research Seminar
March 1, 2012
Modern “Lit” Exchanges

- Fully automated, transparent, real-time order book
- Continuous double auction between buyers/sellers
- Replacing manual/floor exchanges, specialists, etc.
- Many advantages and applications:
  - transparency
  - data-driven algorithmic trading
  - estimating market impact

- Major disadvantage: executing very large orders
  - distributing over time and venues insufficient
  - many buy-side parties are “compelled”
- Thus the advent of... Dark Pools
  - specify side and volume only
  - no price specified, execution by time priority
  - price generally pegged to light midpoint
  - not seeking price improvement, just execution
  - only learn (partial) fill for your order
SEC Weighs New Regulations for Dark Pools

By SARAH N. LYNCH

WASHINGTON – The Securities and Exchange Commission unanimously agreed Wednesday to consider three proposals aimed at shedding more light on non-public electronic trading entities including dark pools, which match big stock orders privately.

The proposals would require dark pools to make information about an investor's interest in buying or selling a stock available to the public instead of only sharing it with a select group operating with a dark pool. They would also require dark pools to publicly identify if their pool executes a trade.

"We should never underestimate or take for granted the wide spectrum of benefits that come from transparency," SEC Chairman Mary Schapiro said. "Transparency plays a vital role in promoting public confidence in the honesty and integrity of financial markets."

Dark pools, a type of alternative trading system that doesn't display quotes to the public, are just one part of a broader probe the SEC is conducting into market structures. Recently, the SEC also voted to consider banning flash orders, which let some traders get a sneak peek at market events. The agency is also looking into other aspects of the market.
The Dark Pool (Allocation) Problem

- Given a sequence or distribution of “client” or parent orders, how should we distribute the desired volumes over a large number of dark pools?
  - a.k.a. Smart Order Routing (SOR), dispersion, etc.
- May initially know little about relative quality/properties of pools
  - may be specific to stock, volatility, volume,…
  - …a learning problem

- To simplify things, will generally assume:
  - client orders all on one side (e.g. selling)
  - client orders come i.i.d. from a fixed distribution
    - …even though our “child” submissions to pools will not be i.i.d.
  - statistical properties of a given pool are static
- All can be relaxed in various ways
- Main contributions:
  - a theoretical framework, algorithm and analysis
  - some empirical validation
Theoretical Framework and Algorithm
Modeling Available Volume: Single Venue

- $v$ shares submitted
- draw $s \sim P$
- execute $\min(v, s)$
- *censored* observations

Graph:
- Probability axis
- Shares available $s$ axis
- Function $P[s]$
Multiple Venues

Client volume $V$ ($V \sim \text{dist. } Q$)

Allocate... ...How?
Two Subproblems

- Optimal allocation under known distributions:
  - greedy algorithm for one-step max fill; other objectives
- Estimating distributions from censored data:
  - Kaplan-Meier is MLE; need new convergence analysis/rate
The Learning Algorithm

- Initially know *nothing* about the venue distributions
  - must simply start allocating each client order
- For each venue, observe (partial) executions
- From censored data, estimate each distribution
  - using an "optimistic" Kaplan-Meier estimator
- From distribution estimates, compute next allocations
  - using greedy allocation on estimates
- Note: our allocations strongly influence observations
  - exploration-exploitation trade-off

- Main claim: simple allocate/re-estimate loop *rapidly converges to near-optimal allocations*
  - exploration is *implicit*: always optimizing w.r.t. current estimates
  - may or may not "fully" learn/explore distributions
Sketch of Analysis

- Algorithm:
  - initialize estimated distributions $P'_1, P'_2, \ldots, P'_k$
  - repeat:
    - compute greedy optimal allocations to each venue given the $P'_i$
    - use censored data to re-estimate $P'_i$ using optimistic K-M

- Analysis:
  - Define “known prefix” $c[i]$ for each $P[i]$
  - if allocation to every venue $i$ is $< c[i]$, already near-optimal
    - know “enough” about the $P_i$ to make this allocation (“exploit”)
  - if for some venue $j$, submitted volume $> c[j]$, we “explore”
    - so eventually $c[j]$ will increase $\rightarrow$ improve $P'_j$
  - optimistic K-M: tail modification ensures always exploit/explore
  - Main Theorem: algorithm efficiently converges to near-optimal
    - non-parametric and parametric versions
Some Empirical Validation
Experimental Framework

- **The Data:**
  - submissions and fills for 12 liquid names x 4 dark pools = 48 pairs
  - proprietary trading flow of large brokerage (internal “clients“)
  - pools: BIDS, AUTO, DE Shaw, NYFIX
  - ~1200 orders, ~1.3M shares per name/pool pair (30-day period)
  - ~16% partial executions, ~9% filled by volume, ~11% censored
  - data cannot be directly used to evaluate algorithms/policies
  - instead use data to build a *parametric simulation* framework

- **The Players:**
  - our allocate/re-estimate algorithm
  - a “bandit“-style allocation algorithm
    - simple weight per venue;
    - multiplicative updates on partial/no fill bit
  - uniform allocation (non-adaptive strawman)
  - ideal allocation with known distributions (unrealizable in practice)
Our Algorithm vs. Uniform Allocation

Fraction Executed

Order Half-life

- learning vs. uniform allocation
- Small orders: +
- Large orders: □
Our Algorithm vs. Ideal Allocation

Fraction Executed

Order Half-life

learning vs. ideal allocation

learning vs. ideal allocation
Conclusions

- Nice no-regret follow-up: Agarwal, Bartlett, Dama
- Other censored trading problems
- Solution for basic dispersion problem; better to condition:
  - targeted volume
  - targeted horizon
  - lit book pressure, buy/sell imbalance, spread,…
- Further info:
  - www.cis.upenn.edu/~mkearns
  - mkearns@cis.upenn.edu