Iceberg Order Detection

Anton Antonov
antonov@dxfeed.com

Quant Insight 2020

CME Iceberg Order Detection and Prediction

www.tandfonline.com/doi/abs/10.1080/14697688.2020.1813904

- Ahead of print in Quantitative Finance
- Print copies available on request
- Preprint version (older!) at arxiv.org/abs/1909.09495

Literature Overview

- Bouchaud et al., *Trades, Quotes and Prices*
- Price impact, front-running strategies
- Hidden liqudity, iceberg orders:
  - Moro et al. (2009)
  - Hautsch and Huang (2010)
  - Christensen and Woodmansey (2013)
  - Frey and Sandås (2017)
  - Fleming et al. (2018)
- C&W (2013): similar framework, predictive model

Problem Statement

Iceberg order $\mathcal{I} = (\mathcal{T}_1, \ldots, \mathcal{T}_M)$, where $\mathcal{T}_r$ is a *tranche*. Then

$$V_{\text{total}} = \text{vol}(\mathcal{I}) = \sum_{r=1}^{M} \text{vol}(\mathcal{T}_r) = \sum_{r=1}^{M} V_{\text{peak}}^{(r)}.$$

Detection

- Identify a sequence of actions $(\mathcal{A}_{r,1}, \mathcal{A}_{r,2}, \ldots)$ that forms $\mathcal{T}_r$.
- Find all $\mathcal{T}_r$ and form a datastructure representing $\mathcal{I}$.
- Compute V_{total}, infer V_{peak}.

Prediction

- Given first $\mathcal{T}_1, \mathcal{T}_2, \ldots$, predict V_{total}.
- Is the iceberg complete (or more tranches will follow)?

Caveats

Tricky:

- $V^{(r)}_{\text{peak}}$ may change
- V_{total} is not, in general, divisible by V_{peak}: $V^{(M)}_{\text{peak}} \leq V^{(r)}_{\text{peak}}$, $r \in \{1, \ldots, M-1\}$
- Hidden volume can be traded directly as the order enters the book

Iceberg Order Types

CME offers native icebergs:

- Order submission $[P, S, V_{\text{peak}}, V_{\text{total}}]$
- Until V_{total} is exhausted, *refill* V_{peak} every time it is fully traded using an order modification message
- Order ID is preserved

Iceberg Order Types

CME offers native icebergs:

- Order submission $[P, S, V_{\text{peak}}, V_{\text{total}}]$
- Until V_{total} is exhausted, *refill* V_{peak} every time it is fully traded using an order modification message
- Order ID is preserved

Independent software vendors offer synthetic icebergs:

- Each tranche is a new limit order, hence no persistent ID: easier to hide, higher costs
- Key assumption: refill happens at the same level P and side S
- No ground truth

Native Icebergs

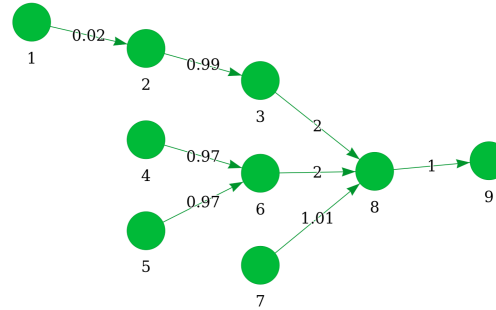
- Finite state machine with states
  - new
  - modify (refill, change price level)
  - initiate trade (as incoming order)
  - affected by trade (as resting order)
  - delete
- FOD LOB historical tapes + CME protocol specification.
- Implemented as an ordered collection of tranches; each tranche is an ordered collection of LOB messages.
- Detection: $\{V_{\text{peak}}, V_{\text{total}}, E\}$, where $E \in \{\text{finished}, \text{cancelled}\}$.

Native Icebergs

- FSA graph (?)

Synthetic Icebergs

- How to link trades and refills?
- C&W (2013): Δt is the maximum allowed time for refill
- Tranche tree and weighting scheme are introduced
- Detection: $\{V_{\text{peak}}, V_{\text{total}}, E\}$, where $E \in \{\text{finished, cancelled}\}$
- Two FSAs: individual order, iceberg



Predictive Model

- Input data: $\{V_{\text{peak}}, V_{\text{total}}, E\}$
- Account for order cancellations: for cancelled icebergs only the lower volume boundary is known

Predictive Model

- Input data: $\{V_{\text{peak}}, V_{\text{total}}, E\}$
- Account for order cancellations: for cancelled icebergs only the lower volume boundary is known
- C&W (2013): kernel density estimate
- Survival analysis with right-censored data: estimate the distribution of V_{total} given a V_{peak}
  - Weighted Kaplan-Meier estimator
  - Bayesian model based on near-ignorant Dirichlet process

Detection Results

- ESU19, E-mini S&P 500 futures contract
- Training sample: 3 days (19M messages)
- Descriptive analysis:
  - Native: 3.8% by traded volume, 0.06% by count
  - Synthetic: 3.3%–14% by traded volume
  - Human bias towards round numbers

Prediction Results

- ESU19 data
- Testing sample: 1 day (6M messages)

Prediction Results

- ESU19 data
- Testing sample: 1 day (6M messages)

Classification:

- Native F1: 0.71–0.86
- Synth F1: 0.58–0.70

Summary

Our extensions of framework in C&W (2013):

- Formalization of native and synthetic icebergs
- Formalization of detection procedure on CME FOD LOB data
- Survival analysis accounts for order deletions
- Hidden volume estimates in line with previously reported
- Detection and prediction are compatible with real-time data stream

Possible Improvements

- More coverage: studies across time and assets
- Performance and robustness
- Better models (e.g. semi-parametric relative risk models with covariates)

Different problem statement:

What's the probability that an incoming order has hidden volume given the state of the LOB (e.g. handcrafted features)?

Internal competition result: AUROC = 0.84

Credits & Contacts

- Links
- Thanks
- dxFeed: www.dxfeed.com
- Quant research unit:
  - Dmitry Zotikov: www.linkedin.com/in/zot
  - Anton Antonov: www.linkedin.com/in/tonytonov