Iceberg Order Detection & Prediction

Anton Antonov

antonov@dxfeed.com

Quant Insights 2020

dxFeed

Icebergs
QI 2020

A. Antonov

Paper

Iceberg
Order Types

Native

Synthetic

Predictive
Model

Results

Summary &
Future Work

CME Iceberg Order Detection and Prediction

- Ahead of print in Quantitative Finance
- Print copies available on request
- Preprint (outdated) version on ArXiV

Literature Overview

- Bouchaud et al., *Trades, Quotes and Prices*
- Price impact, front-running strategies
- Hidden liquidity, 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

Iceberg Order Types

CME offers native icebergs:

- Order submission $[P, S, V_{peak}, V_{total}]$
- Until V_{total} is exhausted, *refill* V_{peak} every time it is fully traded
- 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
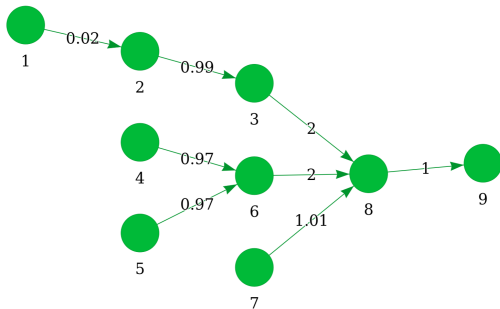- Order ID is preserved

Independent software vendors offer synthetic icebergs:

- 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)
    - trade
    - delete
- FOD LOB historical tapes + CME protocol specification
- Detection: $\{V_{\text{peak}}, V_{\text{total}}, E\}$, where $E \in \{\text{finished}, \text{cancelled}\}$

Synthetic Icebergs

- How to link trades and refills?
- C&W (2013): $\triangle t$ is the maximum allowed time for refill
- Tree structure and weighting scheme are introduced
- Detection: $\{V_{\text{peak}}, V_{\text{total}}, E\}$, where $E \in \{\text{finished, cancelled}\}$

Predictive Model

- Input data: $\{V_{\text{peak}}, V_{\text{total}}, E\}$
- Problem:
  - Infer presence of hidden volume (classification)
  - Infer V_{total} by observing V_{peak} (regression)
  - Account for order cancellations

Predictive Model

- Input data: $\{V_{\text{peak}}, V_{\text{total}}, E\}$
- Problem:
  - Infer presence of hidden volume (classification)
  - Infer V_{total} by observing V_{peak} (regression)
  - Account for order cancellations

- C&W (2013): kernel density estimate
- Survival analysis with right-censored data
  - 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

- Synthetic 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
- Alternative 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

Thanks

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