Generating Realistic Stock Market Order Streams
Source: TEDxNewWallStreet - High frequency trading and the new algorithmic ecosystem by Sean Gourley
Computational Models of Financial Markets

• Analytic Models
  - Extract highly stylized models for tractability to provide key insights;

• Agent-based Models
  - Simulate markets as complex multi-agent systems to reproduce “stylized facts”;

• Generative Models
  - Learn from historical market data to produce realistic and high-fidelity data to facilitate analysis;
Computational Models of Financial Markets

• Analytic Models
  - Extract highly stylized models for tractability to provide key insights;

• Agent-based Models
  - Simulate markets as complex multi-agent systems to reproduce “stylized facts”;

• Generative Models
  - Learn from historical market data to produce realistic and high-fidelity data to facilitate analysis;
  - Generate streams of limit orders that are close in aggregate to the real market of a single stock.
Limit Order Representation

• Limit price
• Quantity
• Type (buy/sell; submit/withdraw)
• Elapsed time
• Best buy price in market
• Best sell price in market
Stock-GAN

\[ L(P_r, P_g) = W(P_r, P_g) = \max_{\omega \in W} \mathbb{E}_{x_i \sim P_r}[f_\omega(x_i|h_i, T_i)] - \mathbb{E}_{z \sim p(z)}[f_\omega(g_\theta(z|h_i, T_i))] \]
Stock-GAN
Evaluation – Synthetic Data

Ground Truth $x_i$

Market Sim 32 ZIs

$g_\theta(z|h_i, T_i)$

$f_\omega$

Critic

History

Time

Generator

Price
Quantity
Type of order
Time since previous order
$\pm 1$ level from order book

$z$

Noise

$h_i$

History

$T_i$

Time
Evaluation – Real Market Data

Ground Truth $x_i$

$\mathbf{ Critic }$

$g_\theta(z|h_i, T_i)$

$\mathbf{ Generator }$

$g_\theta$

$z$

Noise

$h_i$

History

$T_i$

Time

GOOG

$\mathbf{ History }$

$h_i$

$\mathbf{ Time }$

$T_i$
Evaluation – How to Define Realistic?

Ground Truth $x_i$

$f_\omega$

$g_\theta(z|h_i, T_i)$

$z$

Noise

$g_\theta$

$G$

History

$T_i$

Critic

$T_i$

$h_i$

Price

Quantity

Type of order

Time since previous order

$\pm 1$ level from order book

Generator

$G$

History

Time
Evaluation – How to Define Realistic?

• Price - distribution over price for limit orders by order type.
• Quantity - distribution over quantity for limit orders by order type.
• Intensity - number of orders for a consecutive period of time.
• Inter-arrival time - distribution over inter-arrival duration for limit orders by order type.
• Best bid/ask evolution - changes in the best bid and ask over time as new orders arrive.
Preliminary Results

Sim

GOOG

Price Dist.
Preliminary Results

Sim GOOG

Price Dist. Quantity Dist.
Preliminary Results

Sim

GOOG

Price Dist.  Quantity Dist.  Intensity
Preliminary Results

Sim

GOOG

Best Bid and Ask Evolution
Preliminary Results

Sim

GOOG

Best Bid and Ask Evolution

Inter-arrival Time
Conclusion and Discussion

• Demonstrate the effectiveness of using Stock-GAN, a conditional Wasserstein GAN, to generate realistic market data in aggregate;

• Propose to measure the quality of generated limit order streams using five statistics;

• Extend Stock-GAN to generate realistic activity traces of a trading agent, while preserve its trading role or intent.
Thank you! Questions?