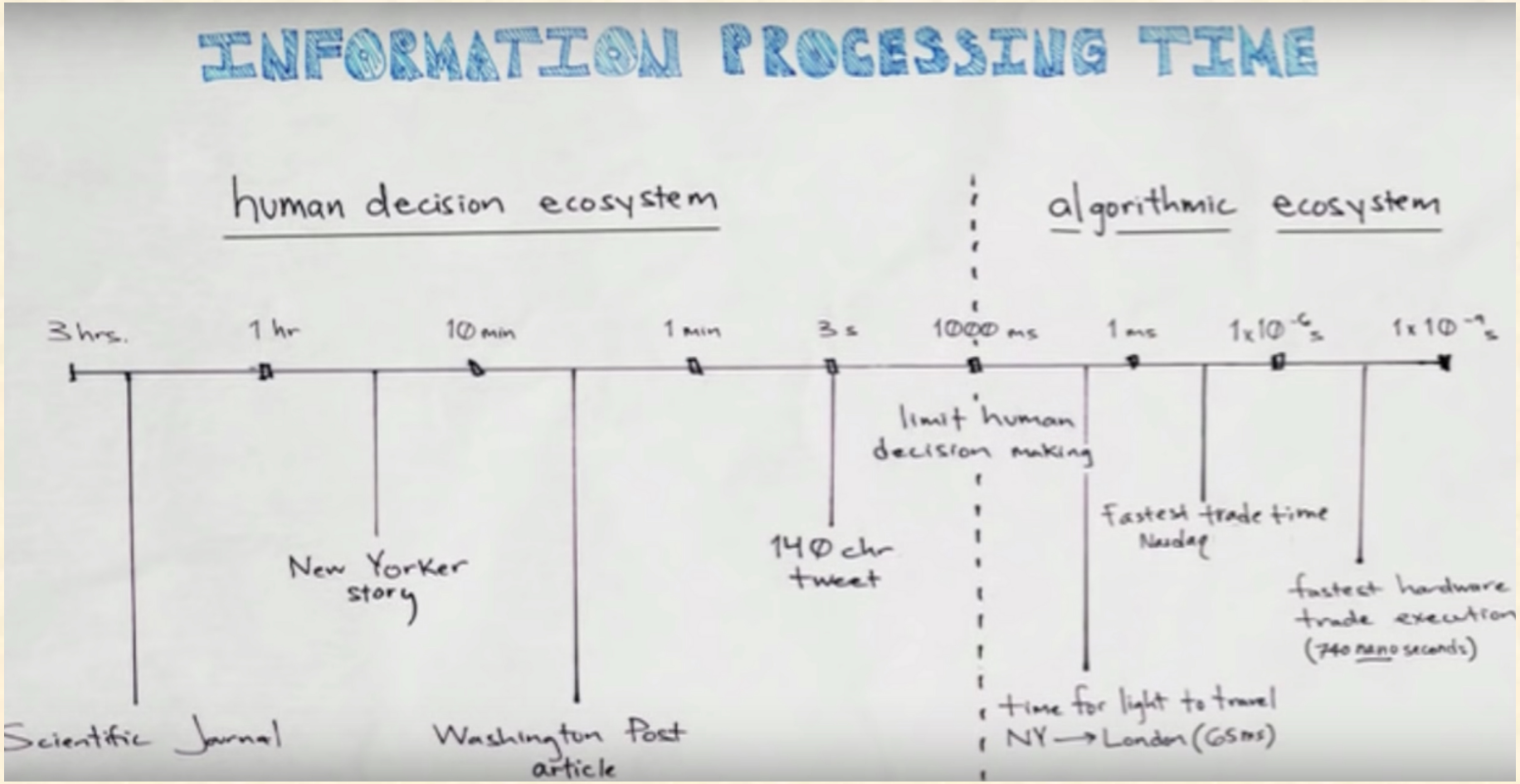
Generating Realistic Stock Market Order Streams

Junyi Li, Xintong Wang, Yaoyang Lin, Arunesh Sinha, Michael P. Wellman

June 2019

Strategic Reasoning Group

University of Michigan



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

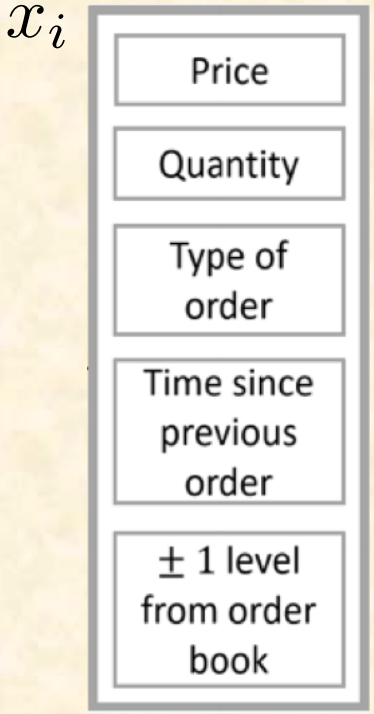
  - Simulate markets as complex multi-agent systems to reproduce “stylized facts”;

- 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.

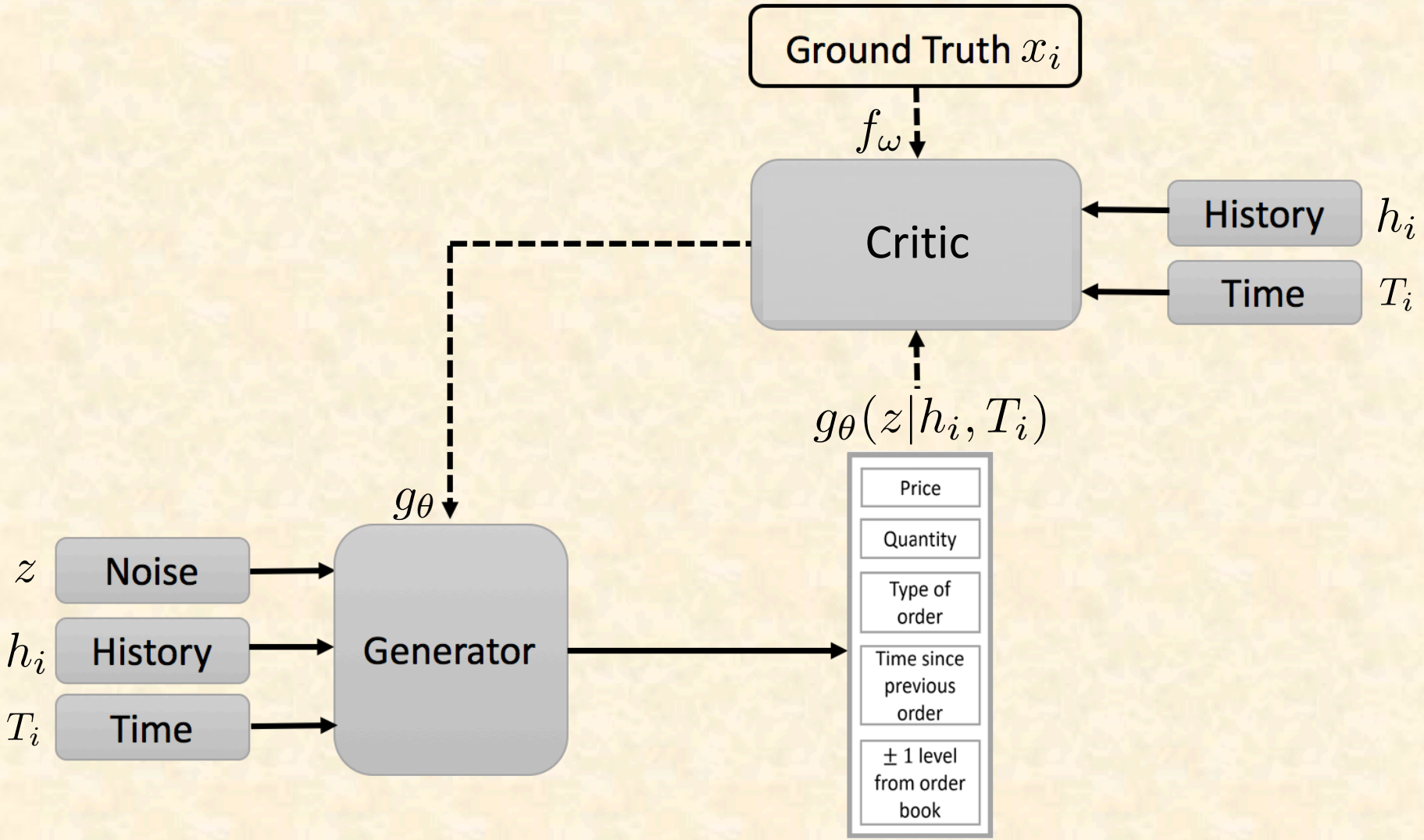
Limit Order Representation

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

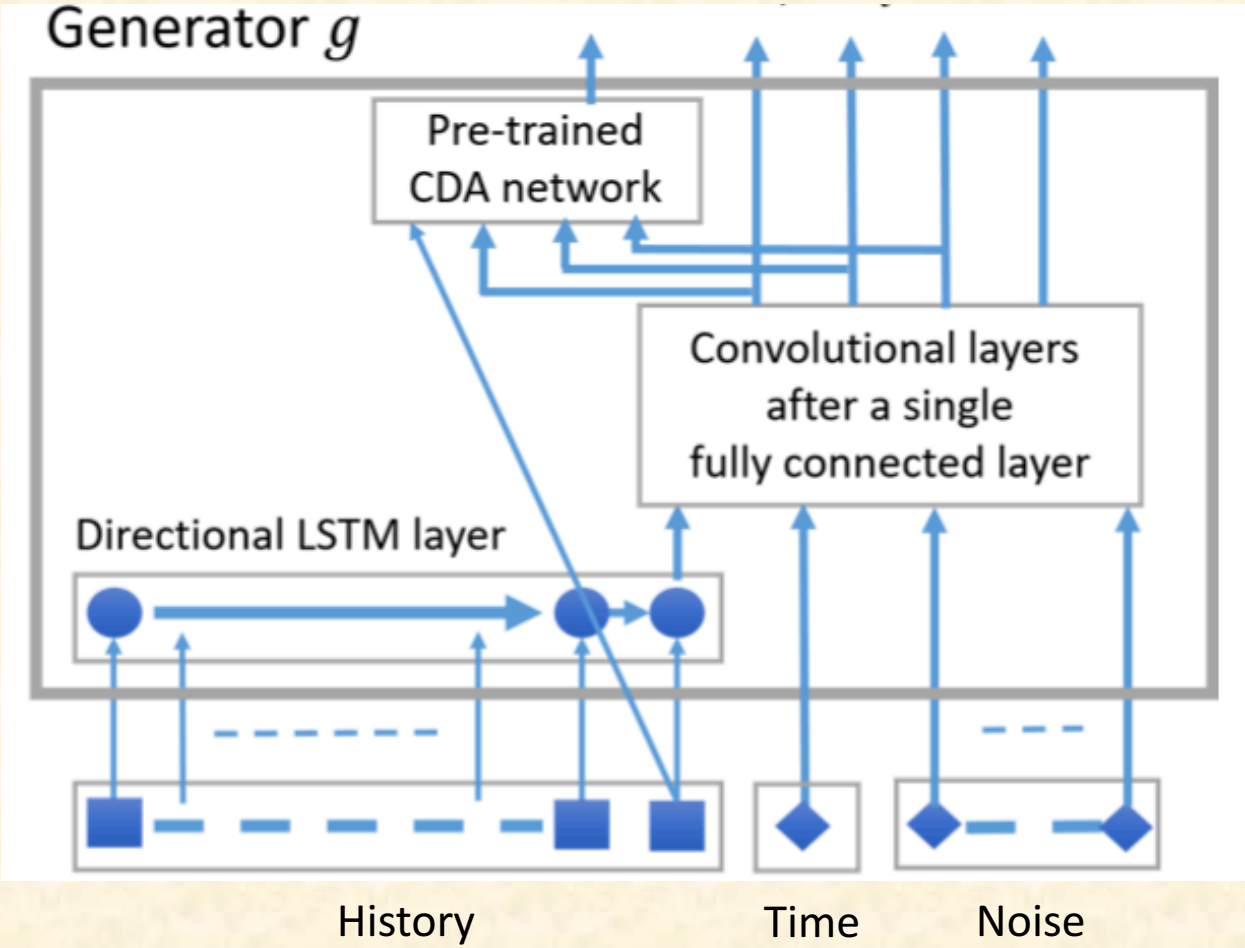


Stock-GAN

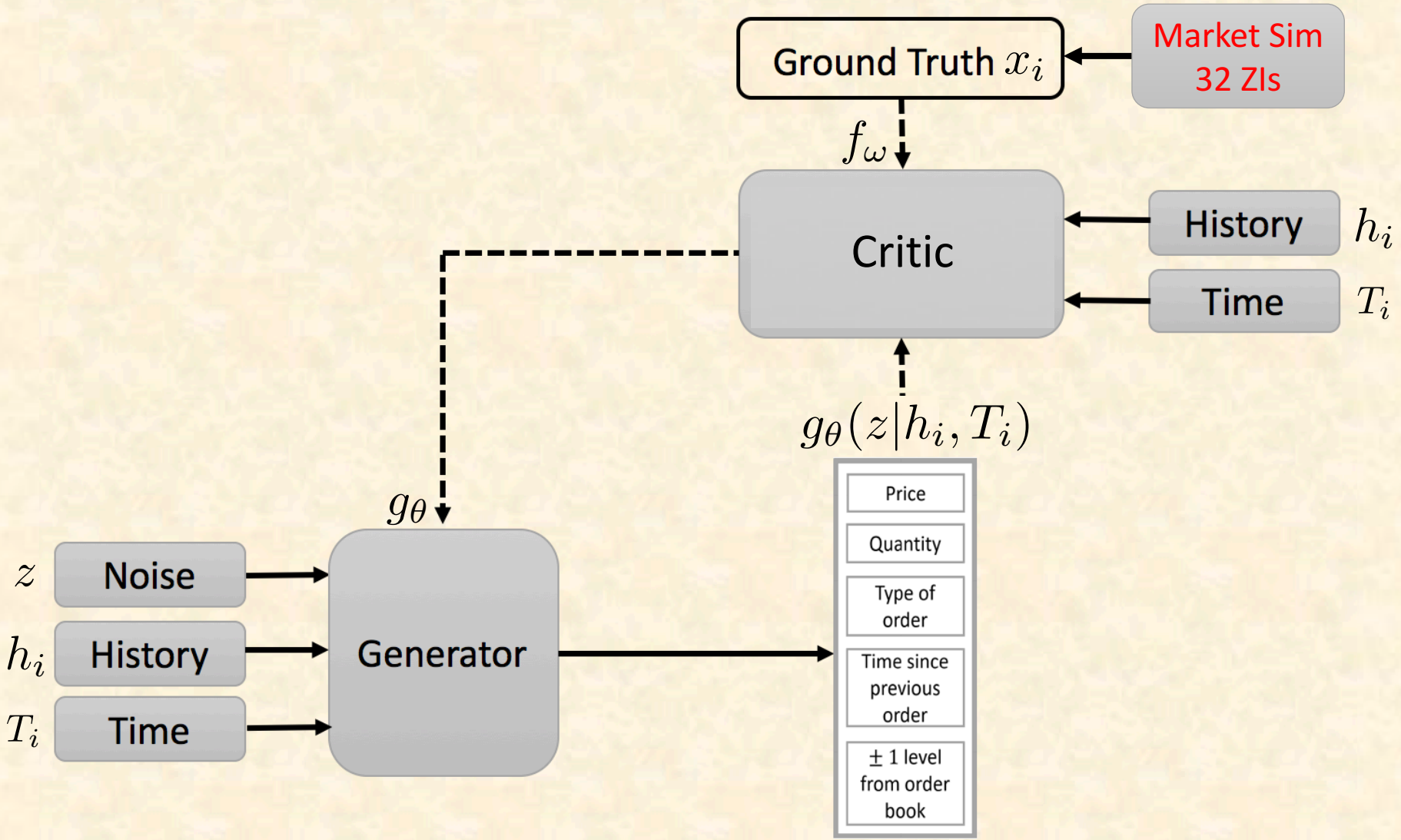
$$L(P_r, P_g) = W(P_r, P_g) = \max_{\omega \in W} \mathbb{E}_{x_i \sim P_r}[f_\omega(x_i|h_i, T_i)] - \mathbb{E}_{z \sim p(z)}[f_\omega(g_\theta(z|h_i, T_i))]$$



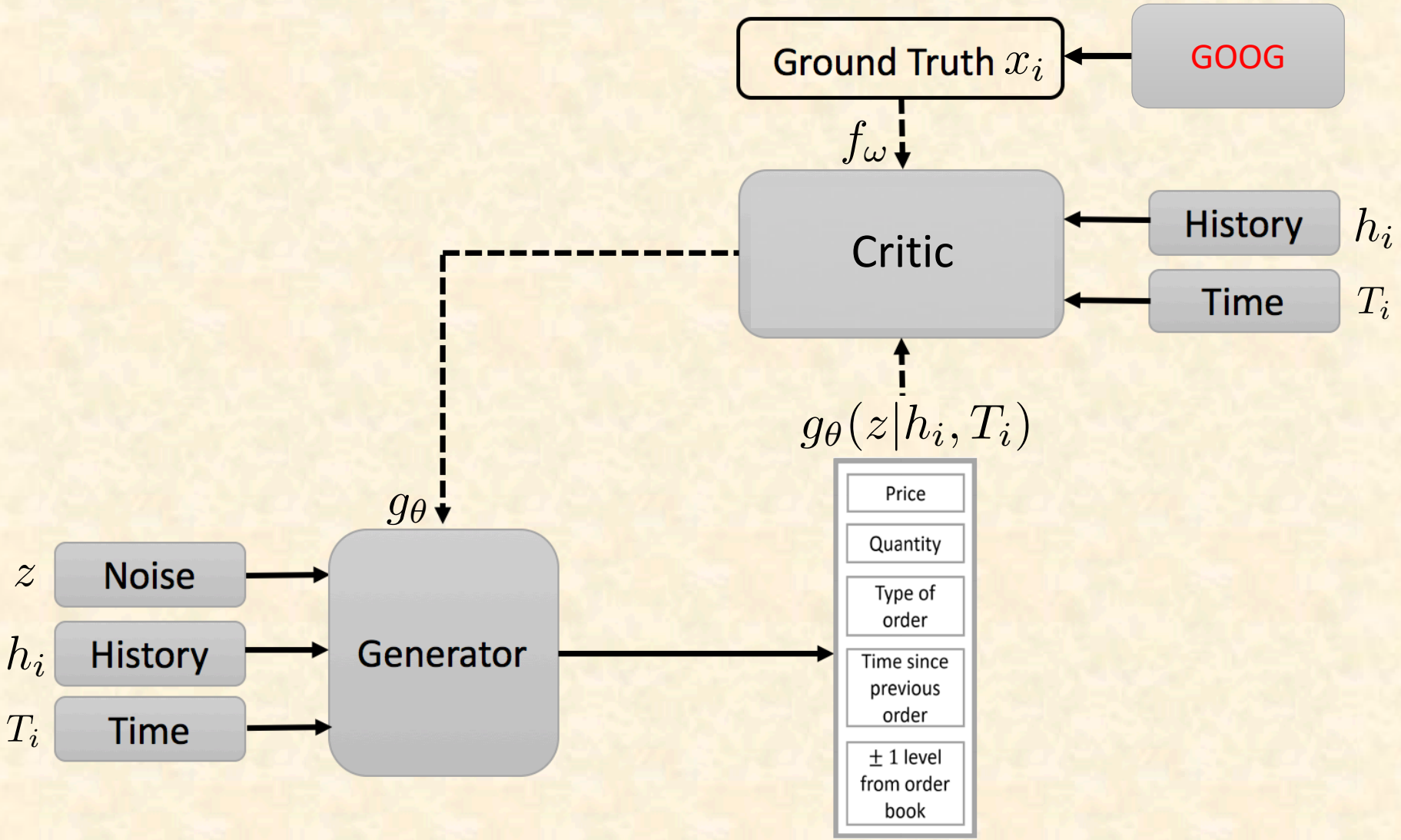
Stock-GAN


Generator g
Pre-trained CDA network
Convolutional layers after a single fully connected layer
Directional LSTM layer
History
Time
Noise

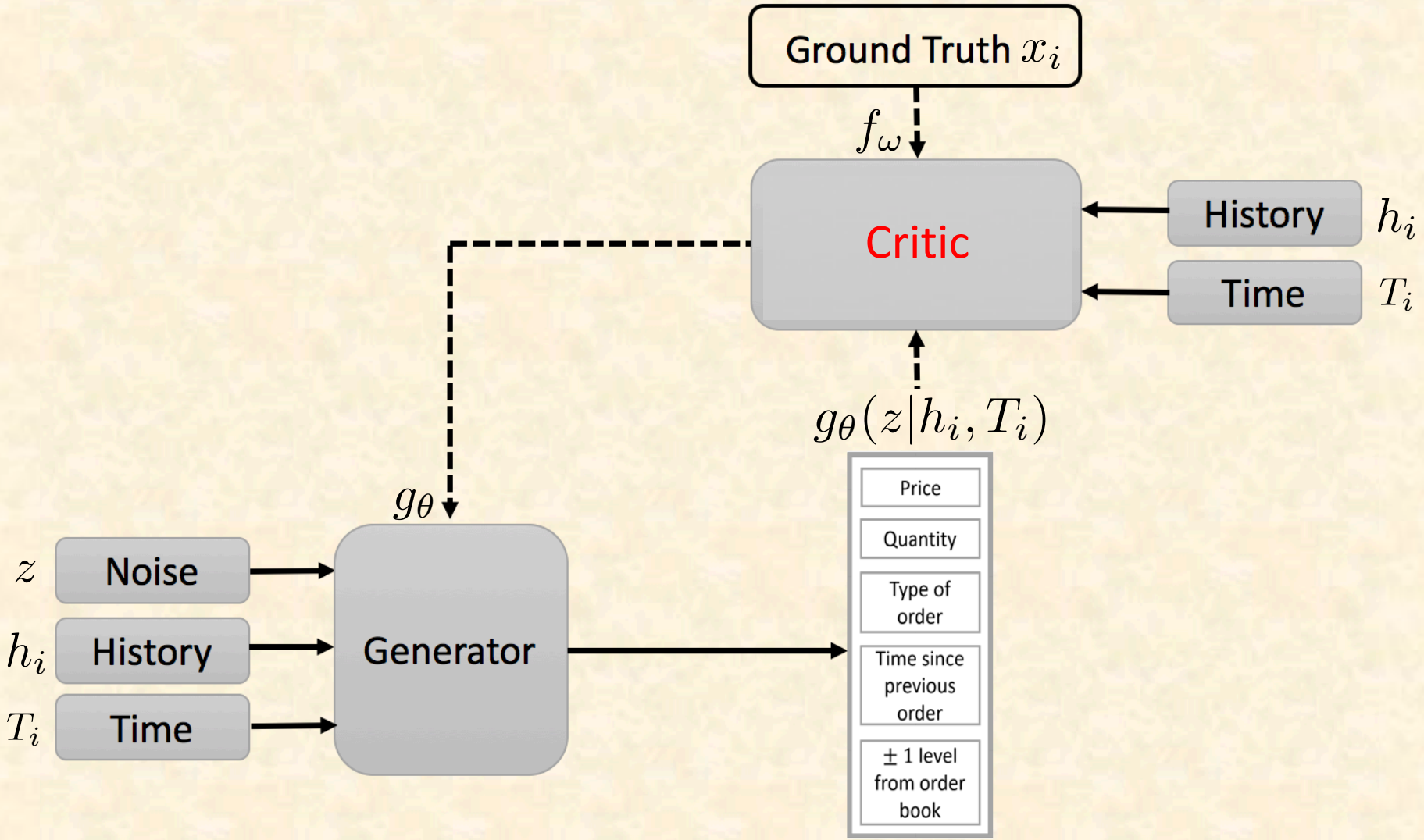
Evaluation – Synthetic Data



Evaluation – Real Market Data



Evaluation – How to Define Realistic?

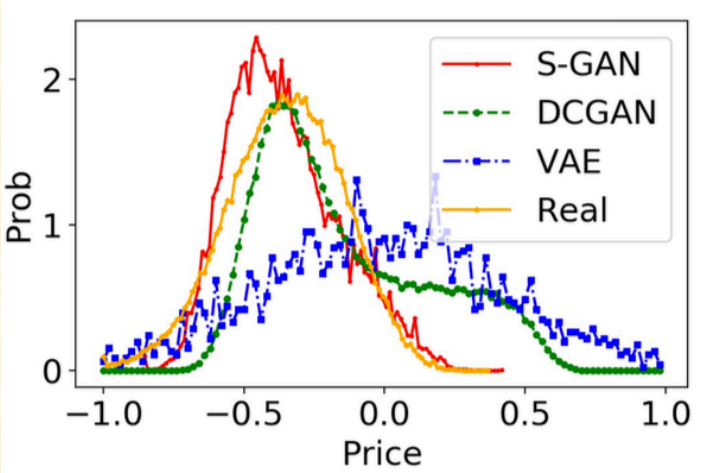


Evaluation – How to Define Realistic?

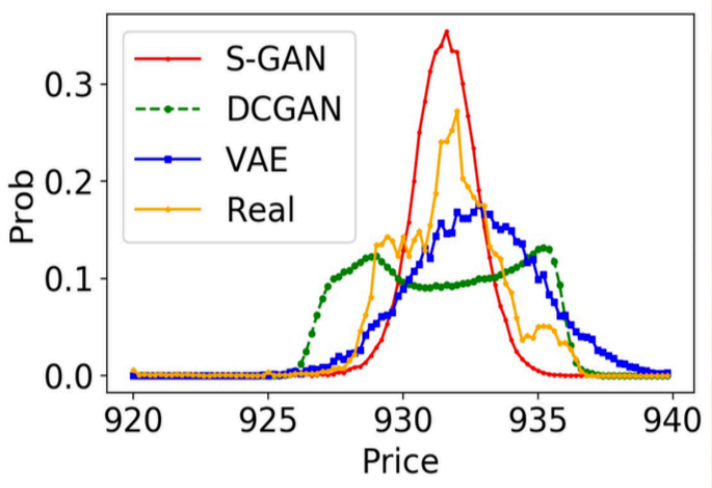
- Price - distribution over price for limit orders by order type.
- Quantity - distribution over quantity for limit orders by order type.
- Intensity - number of orders for a consecutive period of time.
- Inter-arrival time - distribution over inter-arrival duration for limit orders by order type.
-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

Preliminary Results

Sim



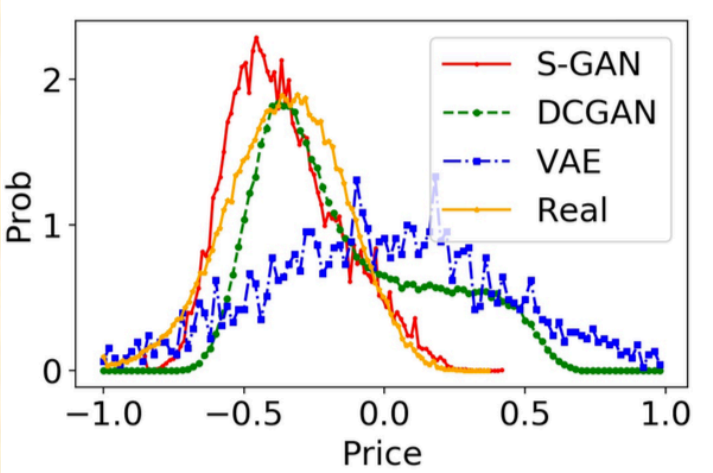
GOOG

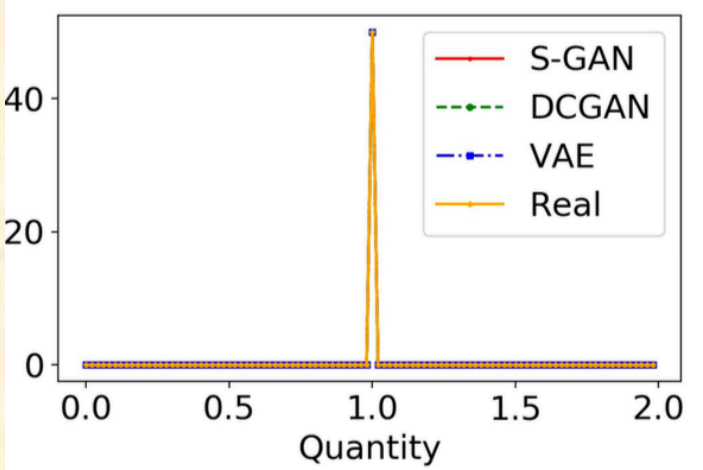


Price Dist.

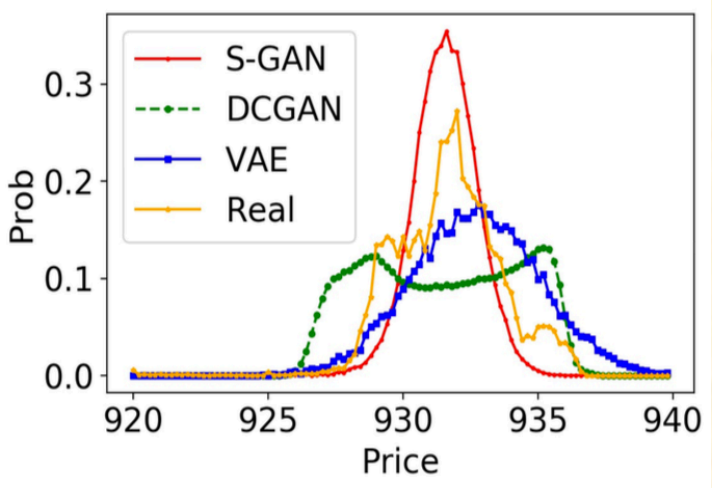
Preliminary Results

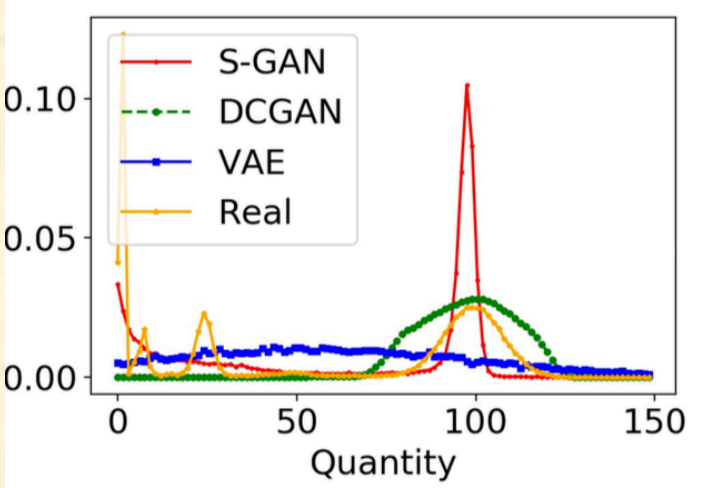
Sim





GOOG



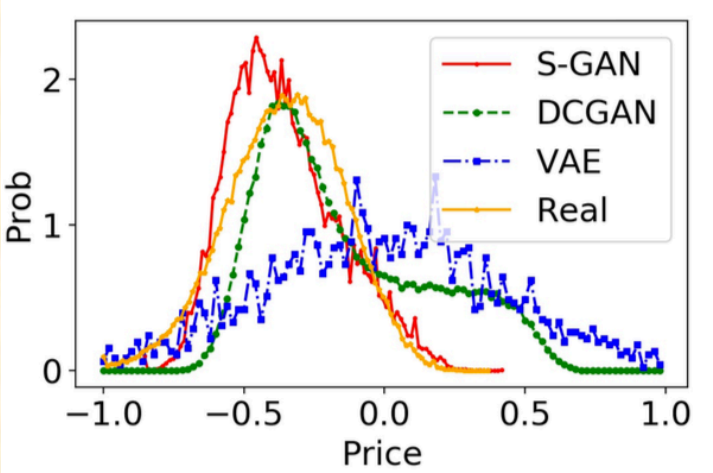


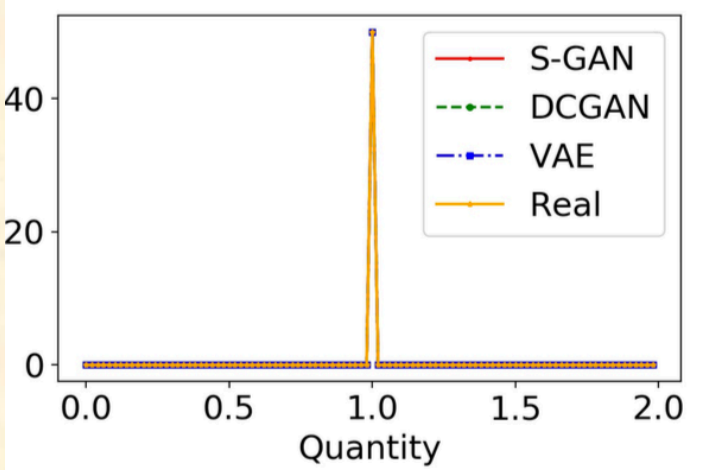
Price Dist.

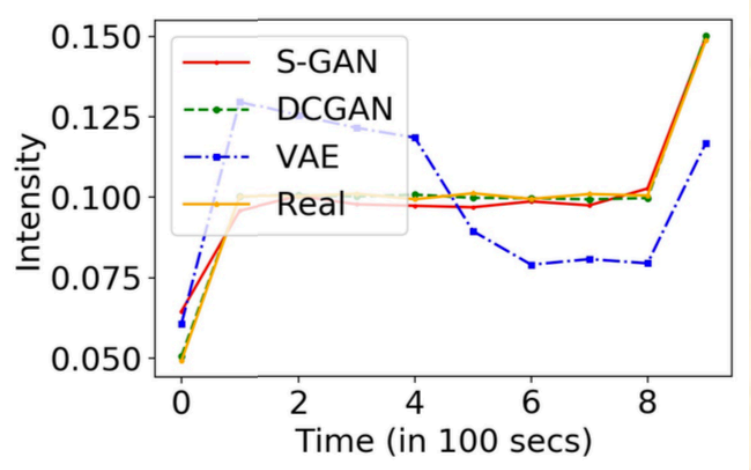
Quantity Dist.

Preliminary Results

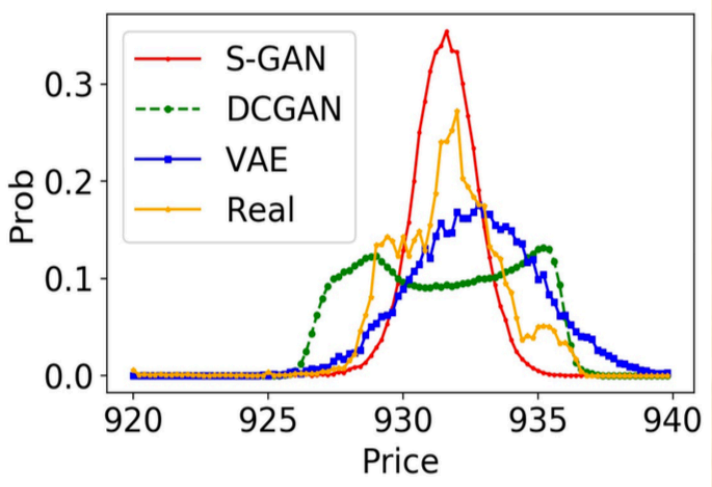
Sim

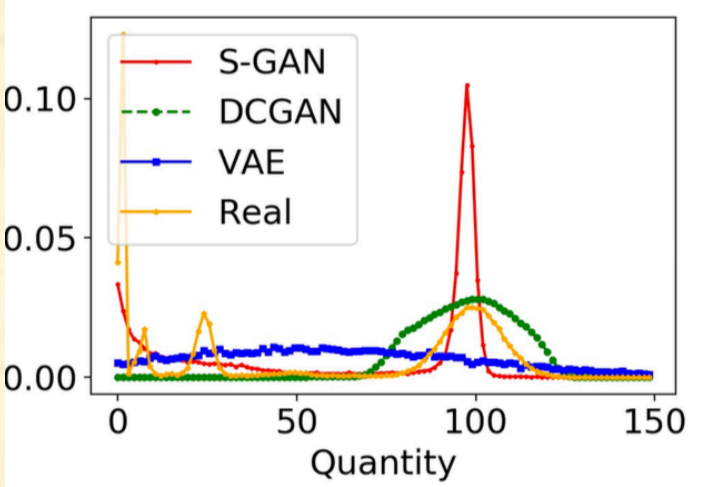


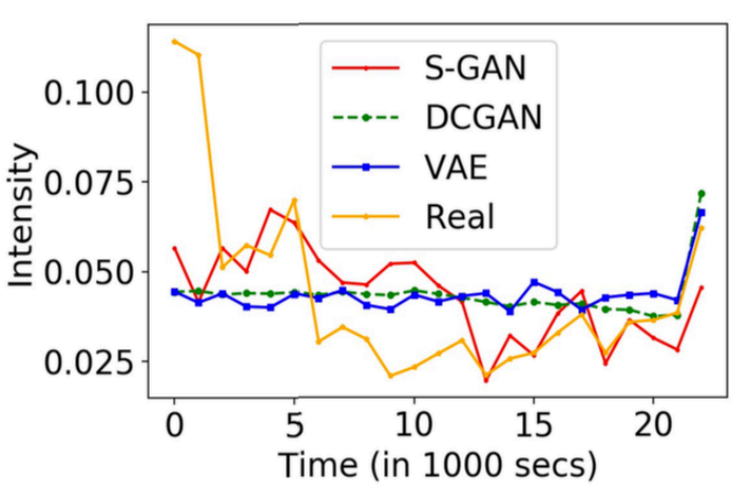




GOOG



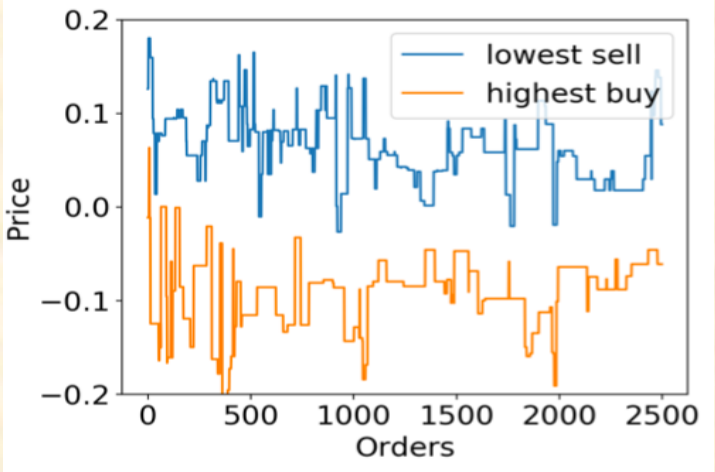


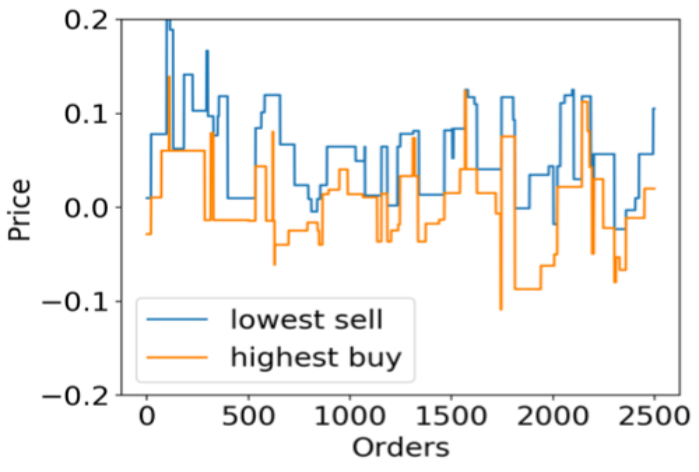


Price Dist. Quantity Dist. Intensity

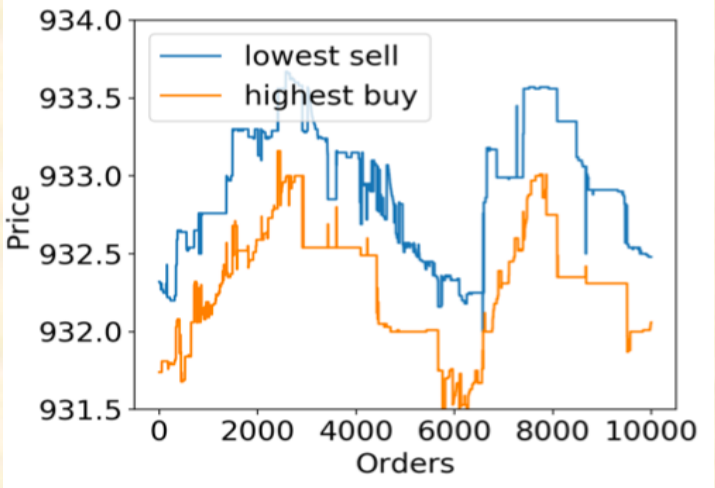
Preliminary Results

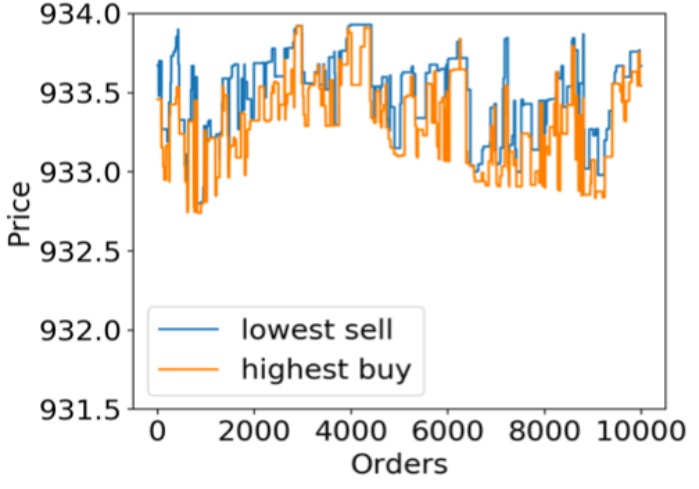
Sim





GOOG

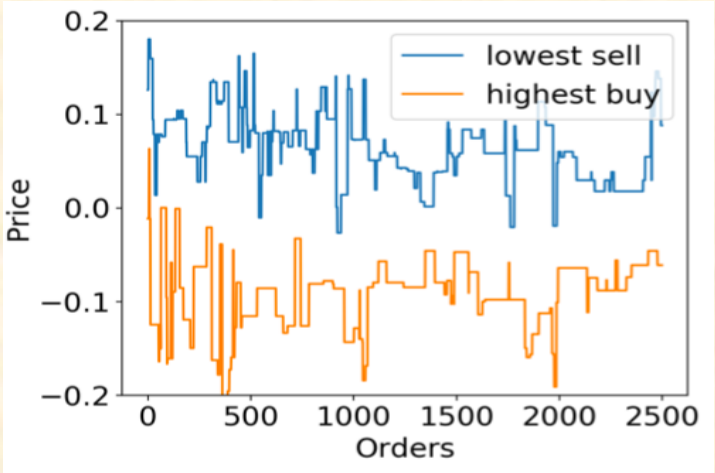


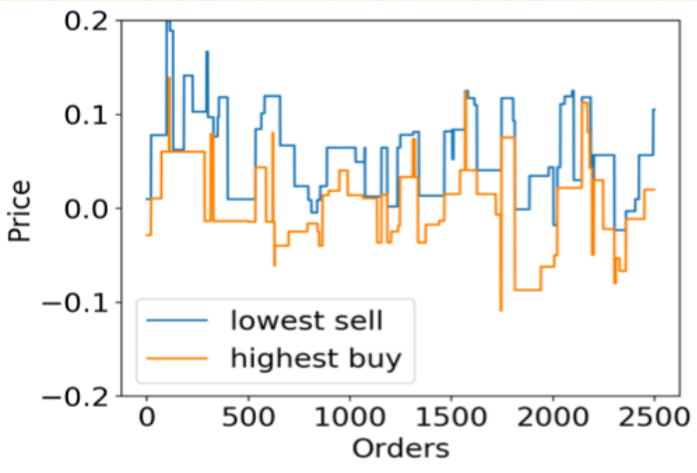


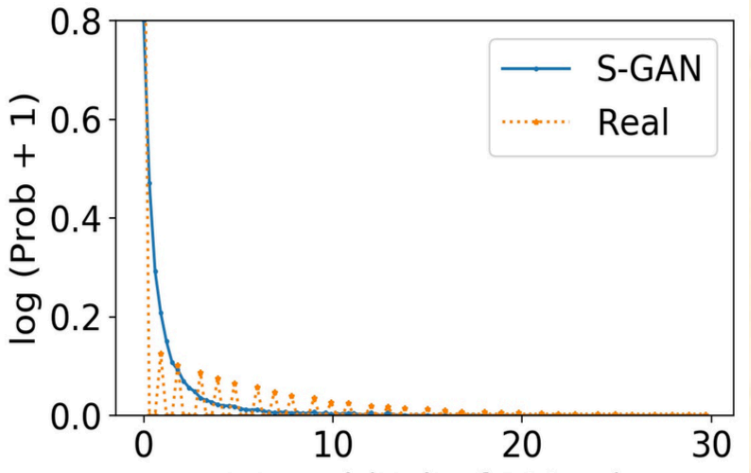
Best Bid and Ask Evolution

Preliminary Results

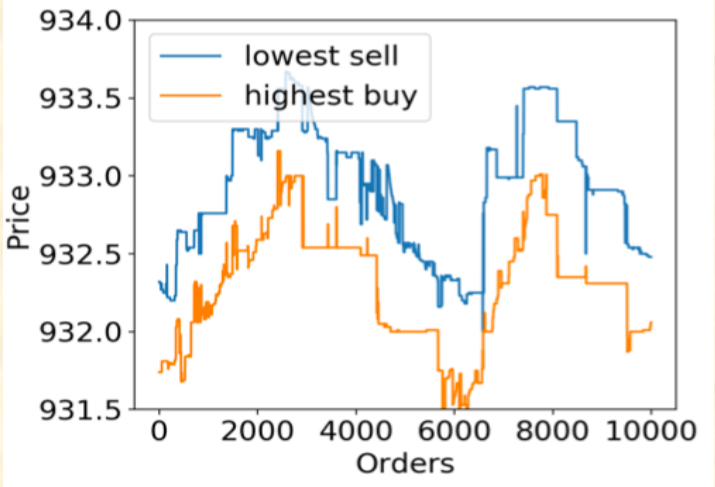
Sim

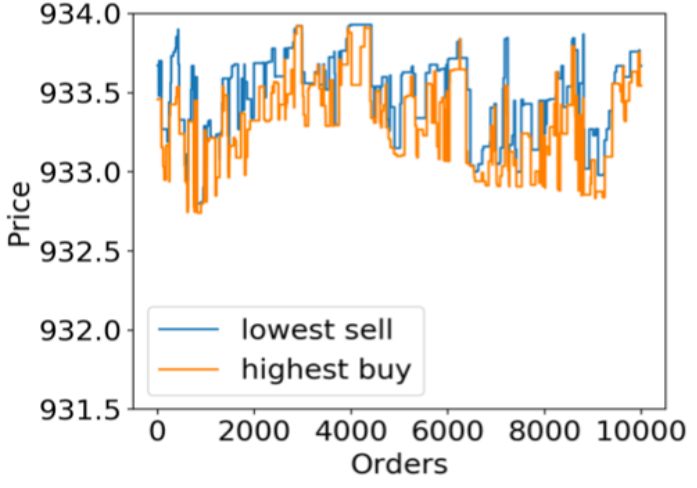


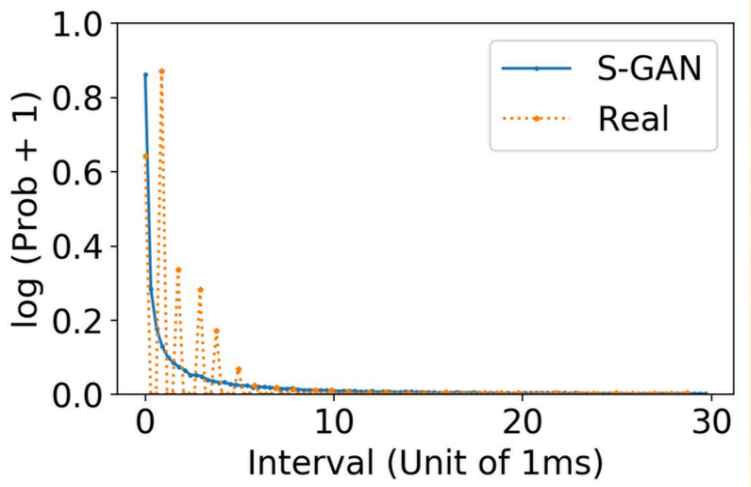




GOOG







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?