Forecasting Prices in the Presence of Hidden Liquidity

Marco Avellaneda,
Josh Reedy & Sasha Stoikov

Replicated by Jamal Verdiyev

Research Questions

- Can Level-I quotes (best bid/ask sizes) forecast short-term price direction?
- Why do naive models fail when visible liquidity is small?
- How does hidden liquidity shape market informativeness?

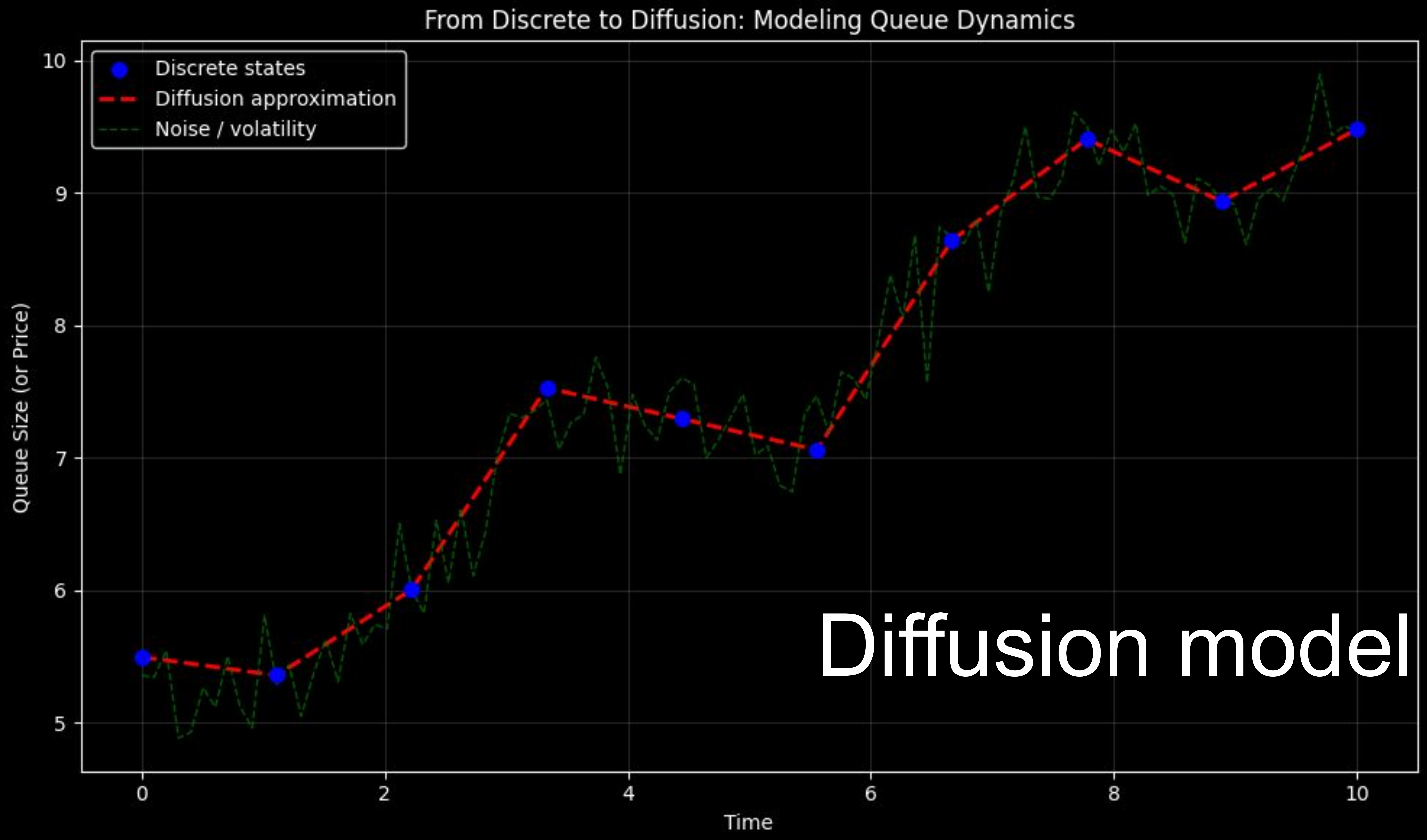
Order Book Basics

- **Level I**: best bid/ask price and size.
- **Level II**: full depth of book.
- Bid/ask queues evolve as orders arrive or are cancelled.
- Key mechanism: price changes when a queue is depleted.

Discrete Poisson Model (CST)

- Queues as Markov chains
- Correlation via diagonal transitions
- But: too complex for practical use

Diffusion model



Mathematical Model: PDE

Diffusion equation for queue dynamics:

$$u_{xx} + 2\rho u_{xy} + u_{yy} = 0$$

Describes the probability that price moves up, given bid/ask sizes (x,y) and correlation ρ .

Derived from functional CLT applied to Poisson-driven queue changes — it's the macroscopic limit of discrete order arrivals.

$$Proposition\ 1. Let\ \Omega(X,Y) - harmonic\ function\colon\ \Omega_{\mathrm{XX}} + \Omega_{\mathrm{YY}} = 0.$$

$$Define\colon v(\xi,\eta) = \Omega\left(\frac{\xi}{\sigma_1},\frac{\eta}{\sigma_2}\right).$$

$$Then\colon \sigma_1^2 v_{\xi\xi} + \sigma_2^2 v_{\eta\eta} = 0.$$

$$Proof\colon$$

$$v_\xi = \frac{1}{\sigma_1}\Omega_X\ ; v_{\xi\xi} = \frac{1}{\sigma_1^2}\Omega_{XX}\ ;\ v_\eta = \frac{1}{\sigma_2}\Omega_Y\ ;\ v_{\eta\eta} = \frac{1}{\sigma_2^2}\Omega_{YY}\ .$$

$$\sigma_1^2 v_{\xi\xi} + \sigma_2^2 v_{\eta\eta} = \Omega_{XX} + \Omega_{YY} = 0\ .$$

$$Proposition\ 2. Let\ \Omega -\ harmonic\ function. Then:$$

$$u(x,y) = \Omega\left(\frac{y+x}{\sqrt{2}\sqrt{1+\rho}}, \frac{y-x}{\sqrt{2}\sqrt{1-\rho}}\right)$$

$$Satisfies\ equation:\ u_{xx} + 2\rho u_{xy} + u_{yy} = 0.$$

$$Proof:\ Let\ \sigma_1 = \sqrt{1+\rho}\,, \qquad \sigma_2 = \sqrt{1-\rho}\,, \qquad \xi = \frac{y+x}{\sqrt{2}}\,, \qquad \eta = \frac{y-x}{\sqrt{2}}\,, \qquad v(\xi,\eta) = \Omega\left(\frac{\xi}{\sigma_1}, \frac{\eta}{\sigma_2}\right).$$

$$Then\ u(x,y) = v\big(\xi(x,y), \eta(x,y)\big), and\ from\ Prop. 1:\ \sigma_1^2 v_{\xi\xi} + \sigma_2^2 v_{\eta\eta} = 0.$$

$$Compute\ derivatives:$$

$$u_x = \frac{1}{\sqrt{2}}\left(v_\xi - v_\eta\right);\ u_y = \frac{1}{\sqrt{2}}\left(v_\xi + v_\eta\right);$$

$$u_{xx} = \frac{1}{2}v_{\xi\xi} - v_{\xi\eta} + \frac{1}{2}v_{\eta\eta};\ u_{yy} = \frac{1}{2}v_{\xi\xi} + v_{\xi\eta} + \frac{1}{2}v_{\eta\eta};\ v_{xy} = \frac{1}{2}v_{\xi\xi} - \frac{1}{2}v_{\eta\eta}.$$

$$Sum\ up:$$

$$u_{xx} + 2\rho u_{xy} + u_{yy} = (1+\rho)v_{\xi\xi} + (1-\rho)v_{\eta\eta} = \sigma_1^2 v_{\xi\xi} + \sigma_2^2 v_{\eta\eta} = 0.$$

Theorem 3.1. Function

$$u(x,y) = \frac{1}{2}\left(1 - \frac{\arctan\left(\sqrt{\frac{1+\rho}{1-\rho}} \cdot \frac{y-x}{y+x}\right)}{\arctan\left(\sqrt{\frac{1+\rho}{1-\rho}}\right)}\right)$$

Satisfies equation $u_{xx} + 2\rho u_{xy} + u_{yy} = 0$, *and* $u(x,0) = 1, \qquad u(0,y) = 0.$

$$X = \frac{y+x}{\sqrt{2}\sqrt{1+\rho}}, \qquad Y = \frac{y-x}{\sqrt{2}\sqrt{1+\rho}}.$$

Proof. *Let* $\Omega(\mathrm{X}, Y) = \arctan\left(\frac{Y}{X}\right)$. *Then* $\Omega -$ harmonic if $\mathrm{X} > 0$. *Apply Prop.* 2:

$$\frac{Y}{X} = \sqrt{\frac{1+\rho}{1-\rho}} \cdot \frac{y-x}{y+x}.$$

$$u(x,y) = \arctan\left(\sqrt{\frac{1+\rho}{1-\rho}} \cdot \frac{y-x}{y+x}\right).$$

$$y \to 0: \frac{y-x}{y+x} \to -1 \Rightarrow \Omega(\mathrm{X},\mathrm{Y}) = \arctan\left(-\sqrt{\frac{1+\rho}{1+\rho}}\right) Consider:$$

$$x \to 0: \frac{y-x}{y+x} \to 1 \Rightarrow \Omega(\mathrm{X},\mathrm{Y}) = \arctan\left(\sqrt{\frac{1+\rho}{1+\rho}}\right)$$

$$But\ we\ need:\ u(x,0) = 1\ and\ u(0,y) = 0.$$

$$Consider:\ f(x,y) = \arctan\left(\alpha \cdot \frac{y-x}{y+x}\right);\ \alpha = \sqrt{\frac{1+\rho}{1-\rho}}$$

$$f(x,0) = -\arctan(\alpha)\ \text{ and }\ f(0,y) = \arctan(\alpha)\,.$$

$$\begin{cases} A \cdot \arctan(\alpha) + B = 0 \\ A \cdot (-\arctan(\alpha)) + B = 1 \end{cases}, and\ if\ follows\ that\ A = -\frac{1}{2\arctan(\alpha)}\ \ and\ B = \frac{1}{2}$$

$$Then:\ u(x,y) = -\frac{1}{2\arctan(\alpha)} \cdot \arctan\left(\alpha \cdot \frac{y-x}{y+x}\right) + \frac{1}{2} = \frac{1}{2} \cdot \left(1 \cdot \frac{\arctan\left(\sqrt{\frac{1+\rho}{1-\rho}} \cdot \frac{y-x}{y+x}\right)}{\arctan\left(\sqrt{\frac{1+\rho}{1-\rho}}\right)}\right)$$

Naive model

$$\text{“}Visible\ liquidity = Total\ liquidity\text{”}$$

$$p(x, y) = \frac{x}{x + y}\ (case\ when\ \rho = -1)$$

Assumptions:

No hidden liquidity.

Price moves up as soon as ask queue hits zero: $u(x, 0) = 1,\ u(0, y) = 0.$

Limitations of Naive Models

- Classical “queue race”: whichever queue hits zero → price moves.
- Predicts: if ask size → 0, then P(up) → 1.
- Reality: empirical probabilities never reach 1.
- Conclusion: visible book alone does not explain moves.

Simplified Probability Formula

$$if\,\rho = -1\; then:$$

$$p(x, y; H) = \frac{x + H}{x + y + H}$$

$x - bid\ size, y - ask\ size, H - hidden\ liquidity$

Interpretation:

$H = 0 \rightarrow$ full transparency.

$H \neq 0 \rightarrow$ sizes lose predictive power.

Hidden Liquidity – obscured part of the market

H = iceberg-orders + liquidity on other exchange

Even if “ask” queue is empty – price may not go up, which ruins the concept of Naive model

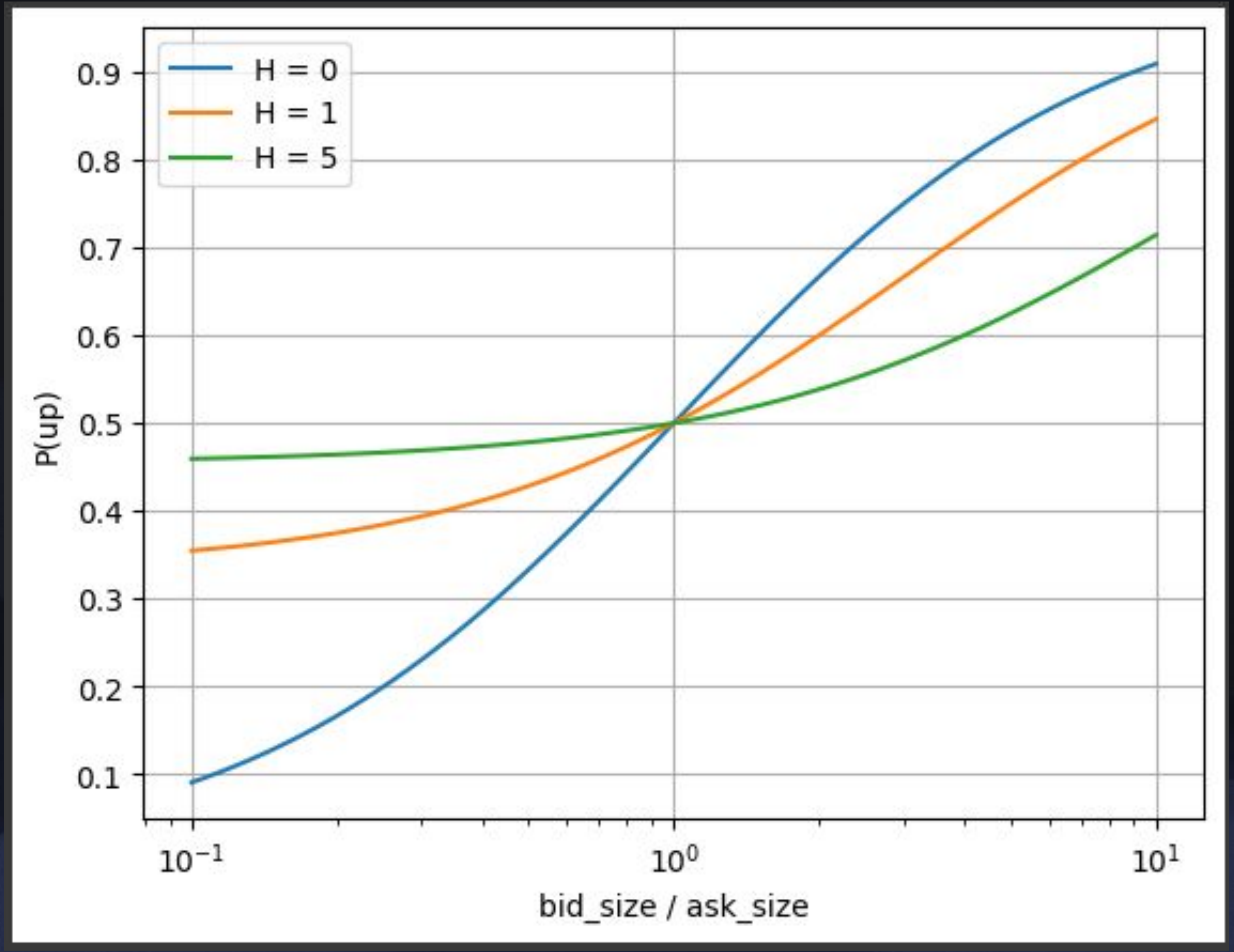
How authors implement Hidden liquidity in the paper?

- True size of the queue now is = visible queue + H

- Probability of price going up is:

$$p(x, y, H) = \frac{x + H}{x + y + 2H}$$

Where x and y – normalized sizes



Data analysis

How authors determine the hidden liquidity on real data?

- Using Level-I data
- Determining for each pair (bid_size, ask_size) where mid-price moves?
- Making deciles -> having buckets (i, j)
- Empirical probability u_{ij}
- Determining H, by minimizing:

$$\sum_{i,j}\left(u_{ij} - \frac{i+H}{i+j+2H}\right)$$

My replication on real data from Lobster

- MSFT, AMZN and AAPL stocks from July 21 2012 – Level-I order book
- Exact replication of authors' algorithm using Python

	bid_decile	ask_decile	up_prob	count
0	1	1	0.532140	8043
1	1	2	0.225257	5931
2	1	3	0.143865	4254
3	1	4	0.149634	3689
4	1	5	0.136515	3604
5	1	6	0.165550	3582
6	1	7	0.188756	3682
7	1	8	0.103534	3226
8	1	9	0.060981	2345
9	1	10	0.034980	2773

	mid	next_mid	direction
0	30.97	30.960	0
1	30.97	30.960	0
2	30.97	30.960	0
3	30.96	30.965	1
4	30.96	30.965	1
5	30.96	30.965	1
6	30.96	30.965	1
7	30.96	30.965	1
8	30.96	30.965	1
9	30.96	30.965	1

Stock	H parameter	Informativeness of Level-I
MSFT	0.000	Very high
AMZN	3.902	Mid
AAPL	8.672	Very low

- **MSFT** (H = 0.000) – as transparent as it gets!
  - When bid_decile=10, ask_decile=1 -> P(up) = 0.97 — almost 100%! This means: if the bid is huge and the ask is small, the price will almost always go up
  - The model fits the data very well

- AMZN (H = 3.902) –moderate informativeness
  - For example, with (8,2) -> P(up) = 0.71, but with (6,2) -> P(up) = 0.75 — almost the same, despite the difference in bid_size.

- AAPL (H = 8.762) –lowest informativeness
  - For example, even (7,1) -> P(up) = 0.56, close to 0.5 despite the difference

Visible quotes ≠ the full picture. Hidden liquidity is determinable and representative on how much you can trust Level-I data.

Fin.