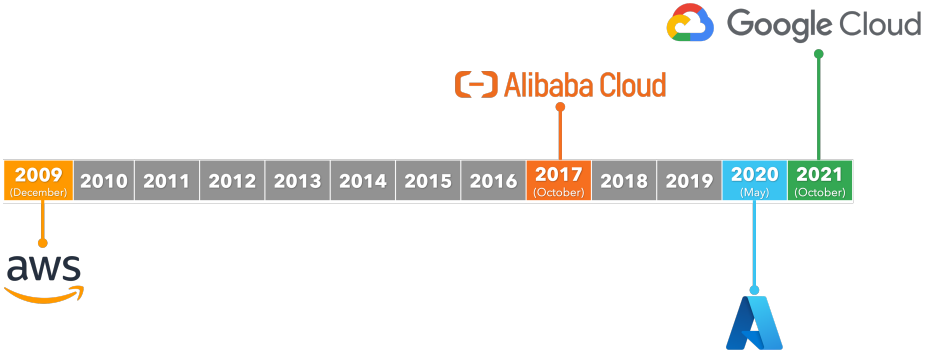
Deconstructing Alibaba Cloud's Preemptible Instance Pricing

Danielle Movsowitz Davidow[1,2] **Orna Agmon Ben-Yehuda**[3] Orr Dunkelman[2]

[1]Department of Industrial Engineering, Tel-Aviv University
[2]Department of Computer Science, University of Haifa
[3]CRI, University of Haifa

Preemptible (Spot) Instances

- Discounted pay-as-you-go instances (compared to On Demand)
- Purchasing:
  - The user specifies a maximum price per hour (bid)
  - If the bid is higher than the market price and the stock is sufficient, the instance is created.
- Announcement History:



Deconstructing Amazon EC2 Spot Instance Pricing, O. Agmon Ben-Yehuda, M. Ben-Yehuda, A. Schuster and D. Tsafrir

Amazon publicizes the spot price but does not disclose how it is determined. They said it was based on supply and demand.

- By analyzing price histories we reverse engineered the pricing algorithm.
- We found prices are usually not market-driven as previously assumed.
- Prices are randomly generated from a tight interval via a dynamic hidden reserve price.
- Our model prevented research errors for clients, other providers and academic researchers, and ...

Spot Price + Paper History

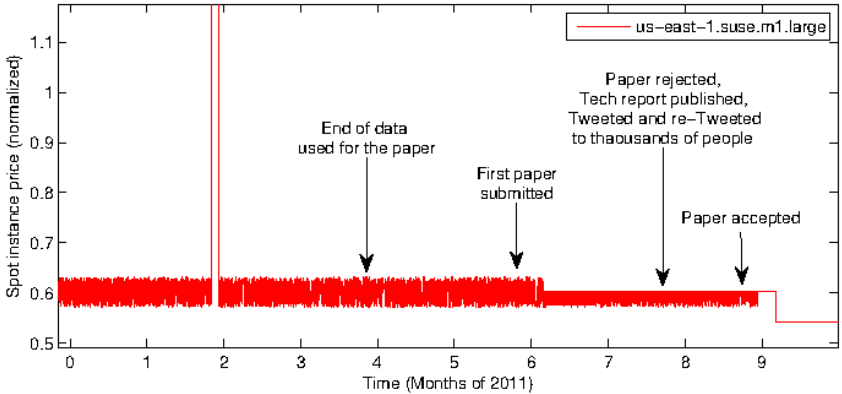
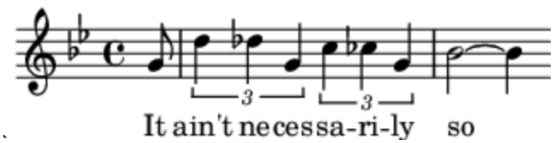


Figure: Price trace with regard to paper publication stages

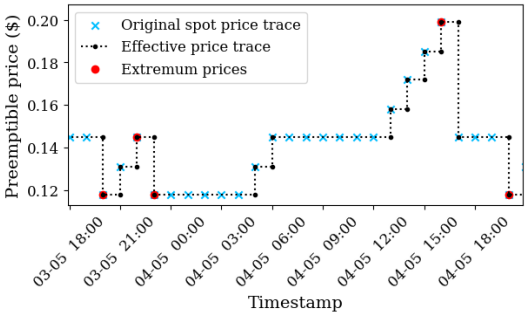
Alibaba Cloud, Supply and Demand

- As of Nov. 2017, AWS “simplified the Amazon EC2 Spot instance pricing by moving to a model that delivers low, predictable prices that **adjust gradually based on long-term trends** in supply and demand”.
- Alibaba Cloud, the largest cloud provider in China, fourth in the world, still states that “The market price of a preemptible instance fluctuates based on changes to the supply and demand of its instance type”.



The Data:
https://crypto.cs.haifa.ac.il/~movso/traces/

- 14,426 trace files (one per instance type X region X availability zone)
- 19 available regions
- November 2018 to July 2021 (many started in February 2020).
- Days of pure missing data: < 40.
- Some more missing data because Alibaba Cloud does not really export 30 days of data.



What goes up must come down...but together?

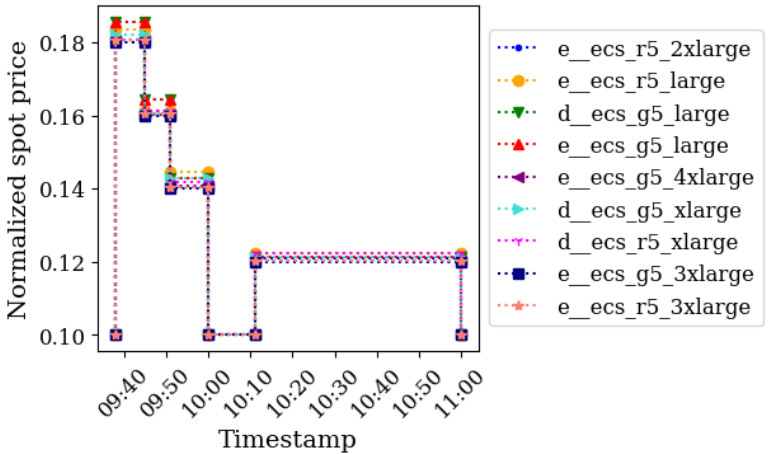
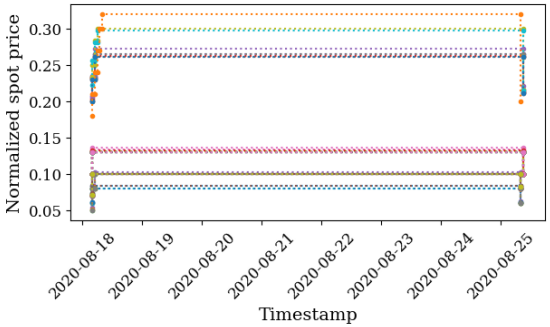
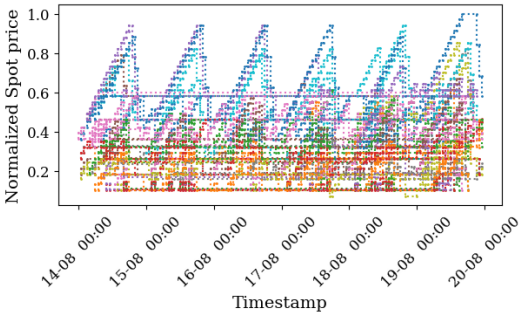


Figure: Instances with correlated normalized prices in region *cn-shenzhen* from zones *e* and *d*.

Synchronization and Correlation





Are these prices set by supply and demand?

Looking for Sharp Behavioral Changes

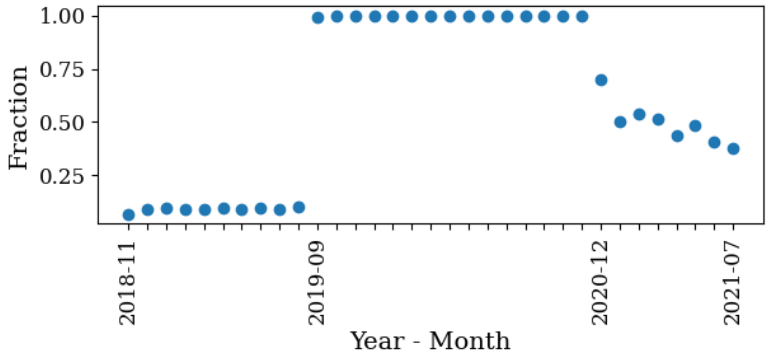
Natural Causes (supply and demand):

- May gradually cause prices to rise or drop
- May gradually change price distribution parameters (e.g., Pareto index after a war)
- Can affect different products differently

Artificial (engineered) Causes:

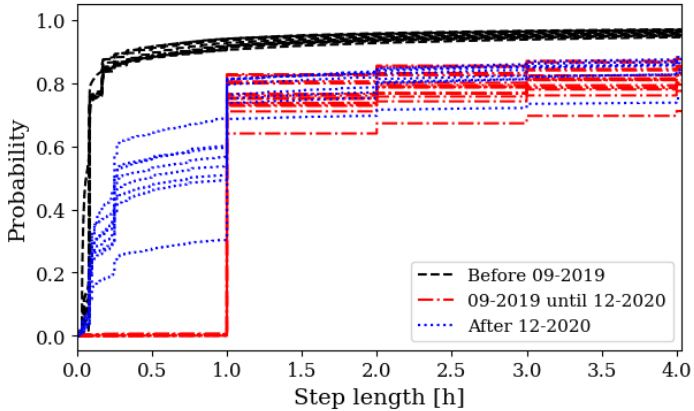
- Can affect all traces at once
- Can change the nature of a distribution
- Might prefer round numbers

Timestamp Analysis: Fraction of price changes at round hours.



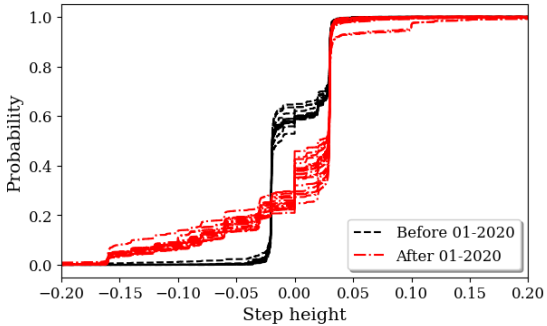
- Sharp changes
- When 100% change at round hours, it is unlikely that price rises are due *supply* $<$ *demand*: If *demand* $>$ *supply* at a non-round hour, the provider must immediately raise the price or dishonor the terms.

Step Length Analysis



Sharp changes in the nature of a distribution.

(Normalized) Step Height Analysis→ The Pacing hypotheses



- A key method: normalize the price by the origin price.
- We are looking at the result of an engineered pricing mechanism, sharply changed on January 2020:
- Alibaba Cloud paces many price raises by dividing them to 3% steps.
- Alibaba Cloud used to pace price drops to 2% decrease steps, but abruptly stopped on January 2020.

Almost, but not exactly, 2% and 3%?

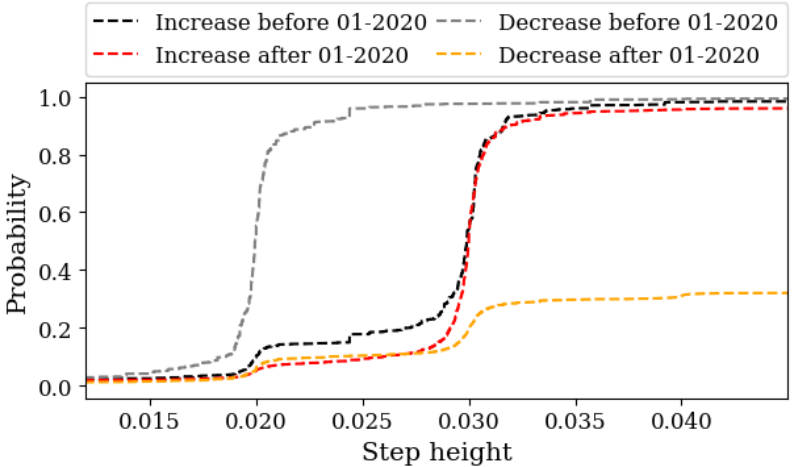


Figure: (Normalized) Step Height Analysis, Folded, by Epochs, Zoom In

Is it due to Quantization Errors?

Bids and prices are expressed with an accuracy of 0.1 cent.
Is this pattern the result of certain normalized steps being impossible to express?

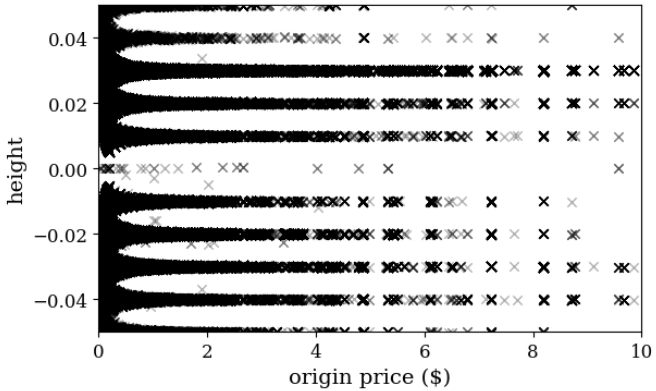
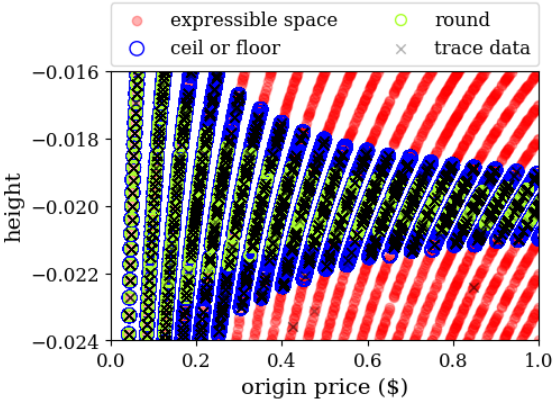


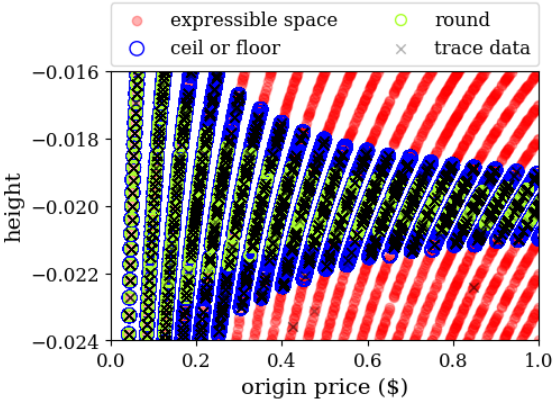
Figure: (Normalized) step height as a function of origin price

Reverse Engineering Decrease Steps



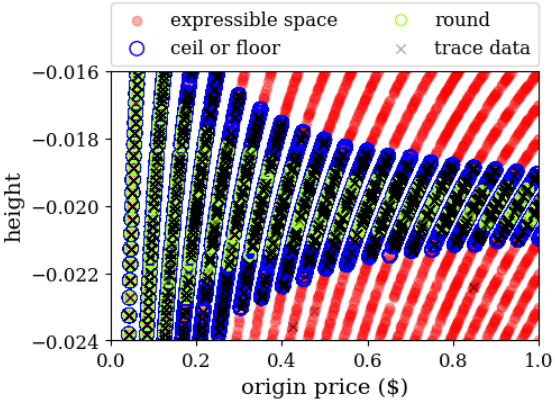
- The expressible space is vast, but rounding covers too narrow a space.

Reverse Engineering Decrease Steps



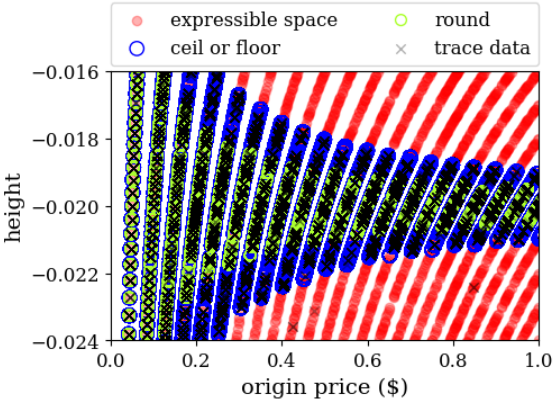
- The expressible space is vast, but rounding covers too narrow a space.
- Randomly using ceil or floor cuts it.

Reverse Engineering Decrease Steps



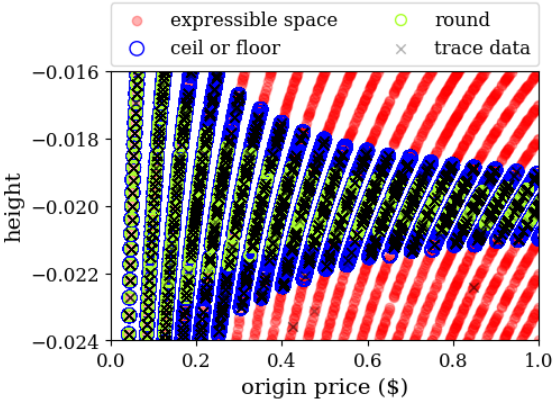
- The expressible space is vast, but rounding covers too narrow a space.
- Randomly using ceil or floor cuts it.
- Alibaba Cloud first chooses a round percent to change the price by,

Reverse Engineering Decrease Steps



- The expressible space is vast, but rounding covers too narrow a space.
- Randomly using ceil or floor cuts it.
- Alibaba Cloud first chooses a round percent to change the price by,
- multiplies it by the origin price and then

Reverse Engineering Decrease Steps

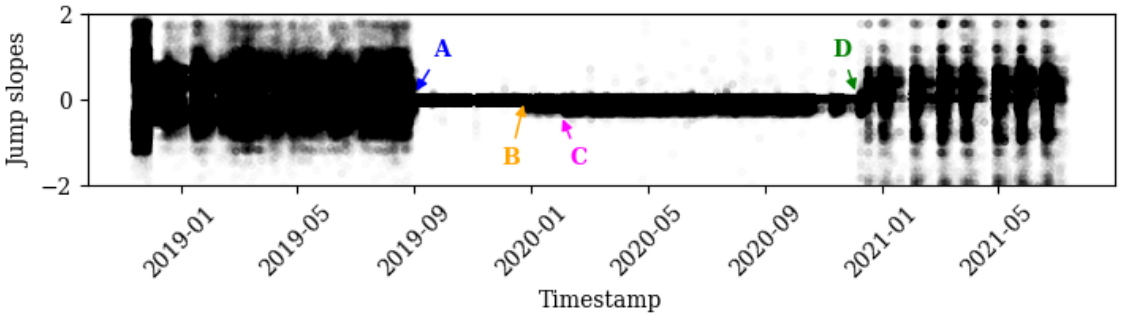


- The expressible space is vast, but rounding covers too narrow a space.
- Randomly using ceil or floor cuts it.
- Alibaba Cloud first chooses a round percent to change the price by,
- multiplies it by the origin price and then
- either uses either floor or ceil, on different occasions. We do not know when it chooses which. Yet.

Jump Analysis: the Quest for the Hidden Variable

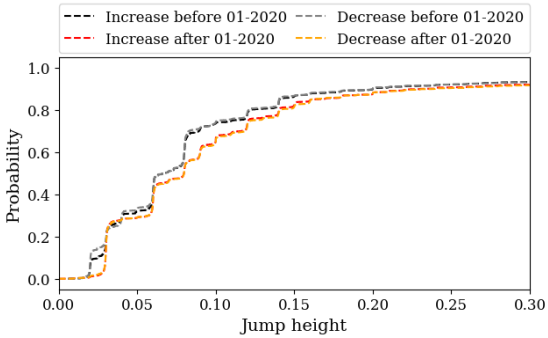
- Are the changes in step height and length related through a hidden variable, that is maintained over sharp changes?
- **Jump**: maximal step sequence in the same direction, with static periods shorter than 48 hours.
- **Slope** (price changing rate) of a normalized jump: its height divided by its length.
- Are the Jump Slope or Jump Height hidden variables?

Jump slopes are not maintained over sharp change events



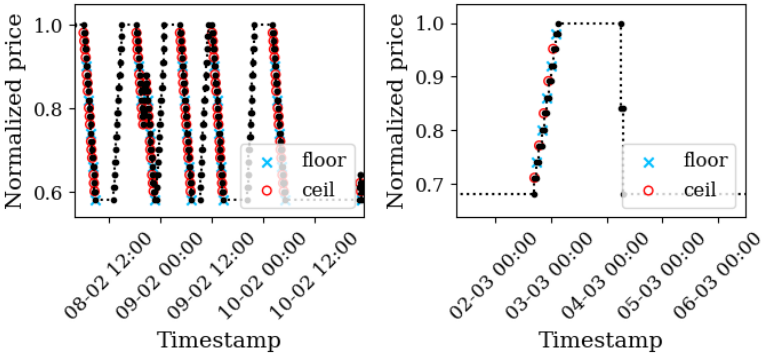
A: change from 5 min intervals to 1 hour intervals between consecutive price changes.
B: change in decrease steps.
C: access to an increased amount of instances.
D: change from 1 hour intervals to shorter intervals between consecutive price changes.

Jump height is the Hidden Variable



The jump height distribution is the same over the behavioral change of January 2020.

How does Alibaba Cloud use Ceil and Floor? Fizz buzz!



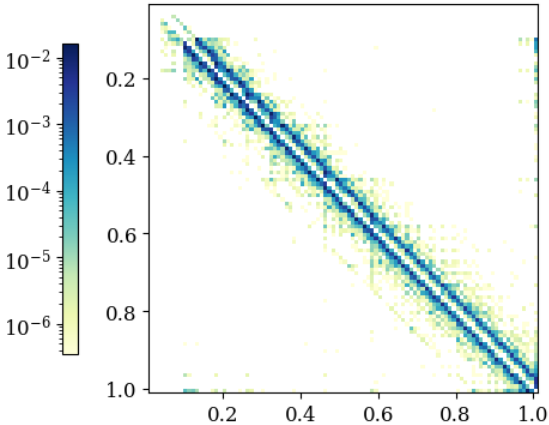
(a) From 2019. (b) From 2021.

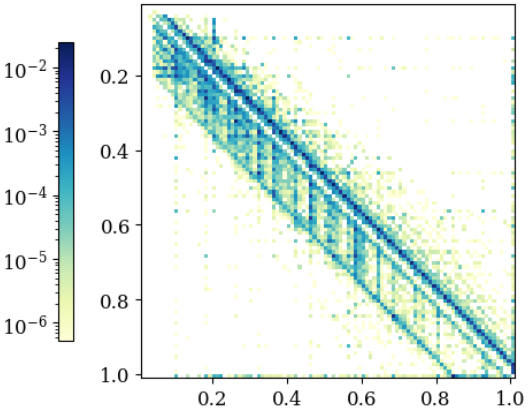
- The slope is approximated along the jump. The last step is usually smaller.
- If the engineered goal was the slope, a small final step would ruin it. We conclude...

Revised Pacing Hypothesis

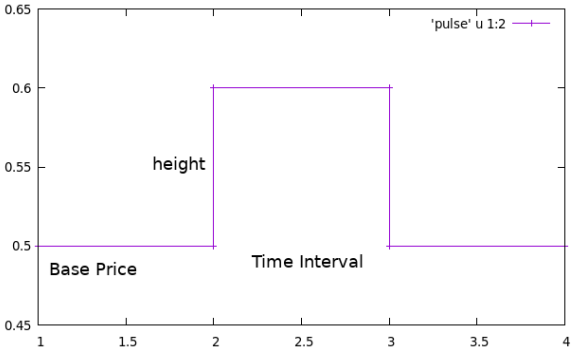
- The jump height distribution is the hidden variable, and is the most important factor.
- Prior to January 2020 decrease jumps were paced at 2% steps.
- Increase jumps are still paced at 3% steps.
- Jumps are engineered according to desired height and length.
- Step sizes are computed using either floor or ceil to maintain a constant slope.
- Any leftover height is mended on the last step.

Price Transition Matrices: Before and After January 2020





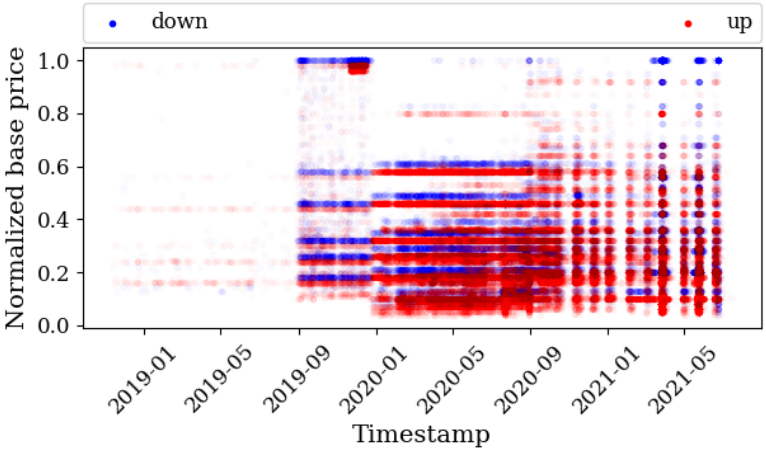
One Hour Pulses



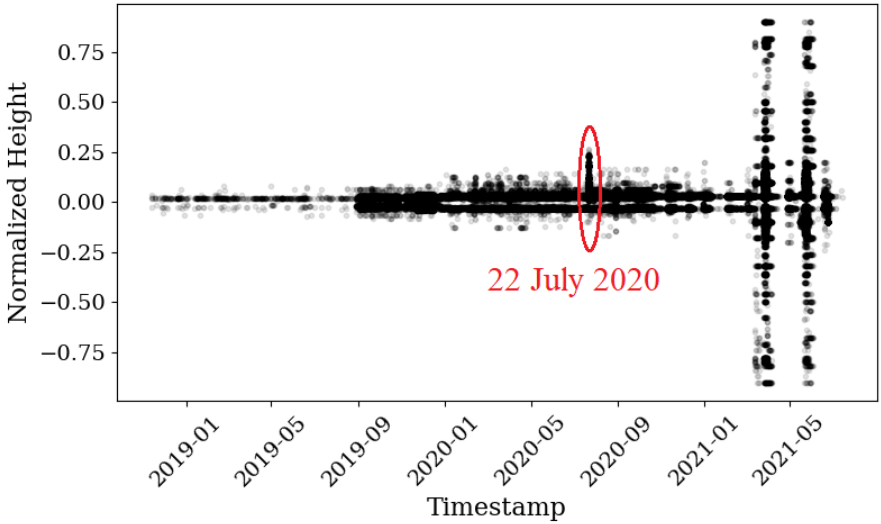
- Pulse patterns appear in 81% of the non-static traces.
- The dominant length is one hour (33%), other lengthes are each at most 6%.
- Hence, we fcused on one-hour pulses.

Normalized base price as a function of the timestamp for all pulses.

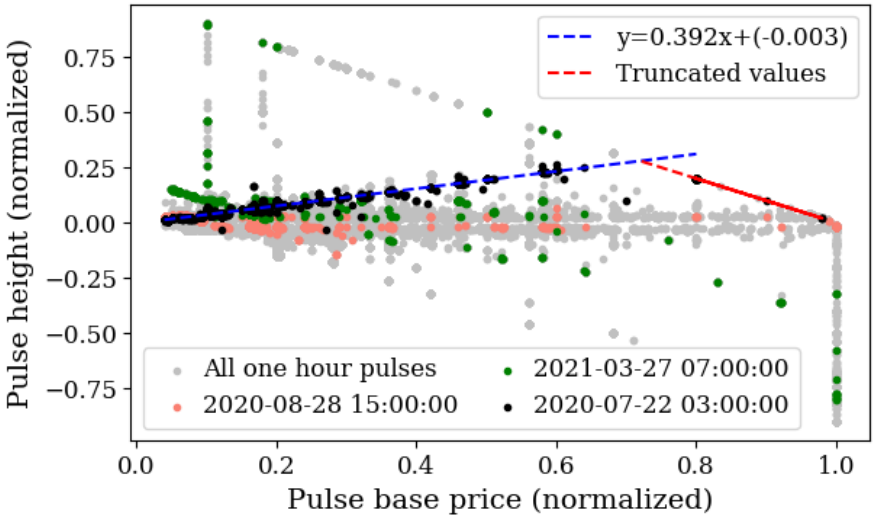
Pulses go up or down, according to the base price.
Sharp changes!



Normalized height of the one-hour pulses as a function of timestamp



July 22nd, 2020, between 3:00 a.m. and 4:00 a.m.



What have we learned?

- A price transition matrix is usefull for price prediction within an era.
- We unveiled sharp changes in artificial trace features, marking ends of eras.
  - 2 Step length distribution changes
  - A change to the jump pacing algorithm
- Jump features are engineered in this order: height, length, slope.
- Price changes are designed as round numbers of percents of their respective origin prices.
- Starting Jan. 2020, there are preferred normalized price levels.
- There is a secret minimal (reserve) price of 4% of the origin price.
- There are laterally correlated events (e.g., July 22^{nd}).

We still have a lot of questions, so -

Any answers?

Orna Agmon Ben-Yehuda ladypine at gmail.com

The traces are available from:

`https://crypto.cs.haifa.ac.il/~movso/traces/`