Oracle Counterpoint: Relationships between On-chain and Off-chain Market Data

Zhimeng Tammy Yang[*], Ariah Klages-Mundt[†], Lewis Gudgeon[‡]

December 2021

We investigate the theoretical and empirical relationships between activity in on-chain markets and the overall pricing and liquidity in off-chain cryptocurrency markets (e.g., {BTC, ETH}/USD price and level of market liquidity). The motivation is to develop methods for proxying off-chain market data using data and computation that is in principle verifiable on-chain. Such methods would provide an alternative approach blockchain price oracles, which relay off-chain data to be accessible by smart contracts but typically rely on some form of trusted party.

We formalize this as the task of finding a function f that maps on-chain observable data to close estimates of off-chain prices. Ideally, a good f will also have two further properties: (i) it is difficult/costly to manipulate the output of f through manipulating the inputs, and (ii) outputs of f are provable on-chain The hypothesis predicating this structure is that off-chain price data (e.g., in USD terms) is incorporated into the behavior of agents in on-chain markets (e.g., mining, block space, and DeFi markets) and that on-chain data thus provides some information about the original off-chain prices.

Features from On-chain Markets We explored relationships in PoW mining, PoS validation, block space markets, network decentralization (e.g., burden on running a full node), usage and monetary velocity, and DeFi liquidity pools and AMMs, including activity on both Bitcoin, Ethereum, and Celo networks. We selected key features from on-chain data based on the literature of fundamental economic models of these markets (e.g., [7, 8, 2, 3, 6, 4]).

For example, [7] models a block space market and finds that the ratio of average demand to capacity $\rho = \frac{\lambda}{\mu K}$ plays an important role in linking users' waiting costs to transaction fees pricing. Here λ is the transaction volume, K is the maximum number of transactions in a block, and μ is the block adding rate. A function emerges, which we'll call $F(\rho)$ that describes the relationship between fee pricing and congestion, which can be translated as

$$\text{tx fees in USD} = (\text{tx fees in ETH}) * \text{price}_{ETH} = F(\rho).$$

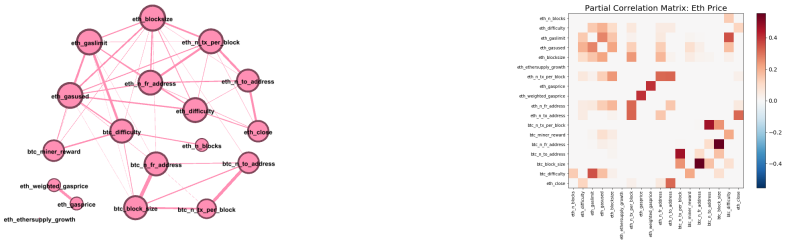
While $F(\rho)$ is nontrivial to work with, various pieces of the results in [7] can be incorporated into useful features for the task of recovering price_{ETH}, including ρ, ρ^2, and the empirical finding that $\rho = 0.8$ represents a phase transition in fee market pricing.

Data Driven Modeling With key on-chain market features in hand, we analyze through graphical models, mutual information, and ensemble machine learning models

[*]Coinbase
[†]Cornell University
[‡]Imperial College London

to explore the degree to which off-chain pricing information can be recovered entirely on-chain.

We use Markov random fields, generated through sparse inverse covariance estimation with graphical lasso regularisation, to express the conditional dependency between the time series of on-chain features and off-chain prices. The output of this technique helps to uncover strong empirical dependencies within the data, suggesting features that are strongly related to price and others that replicate similar information as others. We find that the method is often sensitive to the precise dataset used, which we adjust for by smoothing over the outputs of many k-fold subsets. Figure 1 shows one of the graphical models we find.



(a) Graphical Network Visualisation

(b) Partial Correlation Matrix

Figure 1: Graphical Model Output

We also consider mutual information between features in the dataset, which describes the amount of information (in entropy terms) obtained about price by observing the on-chain features. This is helpful both in identifying strong relationships and evaluating different smoothing factors considering noisy on-chain signals. We find that a large amount of off-chain pricing information is contained in on-chain data.

Finally, we apply various ensemble learning regression models (most notably gradient boosted tree) to the task of recovering off-chain pricing from the on-chain feature set, which we evaluate using out-of-sample test on a rolling basis. The model was trained with a subset of time series data and tested on later data points.

The motivation for using primarily ensemble machine learning models is the non-parametric nature of the dataset and success of ensemble methods in analyzing other market microstructure settings [5]. We find that it is generally hard to recover precise prices in this way except on short time scales with regular retraining (e.g., 1-hour retraining) on feature sets that don't contain DEX pricing of stablecoins. Using stablecoin DEX pricing outright makes the implicit assumption that 1 stablecoin = 1 USD, which is just trusting it as an oracle. Figure 2 shows an example of out-of-sample testing for ETH price recovery in this way, which recovers a lot of price signal, but remains fairly noisy.

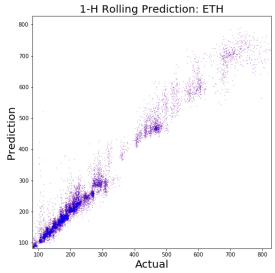


Figure 2: Gradient Boosted Tree Output

Both the noise of these results and the difficulty of replicating hourly retraining models in a provable way on-chain present continuing challenges for this approach.

Proposing an Oracle Consistency Check Off-chain price information can most reliably be recovered from observing on-chain AMM DEXs, as might be expected from work such as [1]. However, recovering {BTC,ETH}/USD prices is not straightforward without making large assumptions about a stablecoin, which involves trusting it as an oracle. We propose a method to use AMM information as a guardrail to inform about the integrity of oracle price feeds without making such large stablecoin assumptions. Our method includes two classes of consistency checks that are designed to add quantifiable costs to manipulating price feeds as a successful manipulation would need to manipulate on-chain time-weighted average prices (TWAPs) in AMM pools.

A first check verifies that several USD-based asset prices that are quoted by an oracle are consistent with the pair prices on large AMM pools (relative, not USD prices). An interesting optimization problem results: using more pools adds security but also adds gas costs and potential DOS vectors, as manipulating the weakest pools can cause the check to fail.

The second check verifies that one highly liquid pair reported by the oracle (e.g., ETH/USD) is consistent with independent oracles (e.g., Coinbase signed prices) and on-chain AMM pools that implicitly use stablecoins as oracles. This is intended to ground the overall price levels, which are not guaranteed by verifying relative prices alone. The second check requires careful filtering (e.g., taking minimums, medians, or quantiles) to increase manipulation and failure resistance. This could potentially be strengthened by additionally checking consistency with ETH price recovery from other on-chain markets, as we explore earlier, should methods for that mature further.

References

[1] Angeris, G., Chitra, T.: Improved price oracles: Constant function market makers. In: Proceedings of the 2nd ACM Conference on Advances in Financial Technologies. p. 80–91 (2020)

[2] Athey, S., Parashkevov, I., Sarukkai, V., Xia, J.: Bitcoin pricing, adoption, and usage. Working Paper No. 3469 (17) (2016)

[3] Buterin, V.: Blockchain Resource Pricing pp. 1–32 (2018), `https://ethresear.ch/uploads/default/original/2X/1/197884012ada193318b67c4b777441e4a1830f49.pdf`

[4] Easley, D., O'Hara, M., Basu, S.: From Mining to Markets: The Evolution of Bitcoin Transaction Fees, SSRN: https://ssrn.com/abstract=3055380 (2018). https://doi.org/10.1007/s10551-015-2769-z.For

[5] Easley, D., López de Prado, M., O'Hara, M., Zhang, Z.: Microstructure in the machine age. The Review of Financial Studies **34**(7), 3316–3363 (2021)

[6] Fanti, G., Kogan, L.: Economics of Proof-of-Stake Payment Systems PRELIMINARY AND INCOMPLETE (November 2018), 1–24 (2019)

[7] Huberman, G., Leshno, J.D., Moallemi, C.: An Economic Analysis of the Bitcoin Payment System*. SSRN Electronic Journal pp. 1–60 (2019)

[8] Prat, J., Benjamin, W.: An Equilibrium Model of the Market for Bitcoin Mining. CesIfo Working Papers (January), 26 pages (2017)