Quantifying The Value Of Order Flow in DeFi: Evolving the DODO

Order Flow Labs
0xD0688E7@protonmail.com

September 17, 2021

Abstract

The cryptocurrency economy has recently seen the dramatic growth of decentralized cryptocurrency exchanges (DEXes) with the Constant Product Market Maker (CPMM) model being by far the most prevalent. Typically these liquidity pools for token pairs have charged a fixed rate fee on swap transactions with a hope that the transaction fee will fully compensate for the risk of impermanent loss. Within this paper we will explore the usage of bayesian analysis to understand the prevalence of informed trading vs uninformed trading and how a simple extension of a newer type of automated market maker could directly address issues around adverse selection.

Background on Automated Market Making

Automated Market Making (AMM) is the umbrella term used to describe several different families of cryptocurrency liquidity providers that operate in an on chain manner. There are numerous benefits to having on chain market making, amongst them are the potential to earn fees for those providing liquidity, less risk of having keys stolen because of a centralized exchange getting hacked and more transparency around fees being charged. Within the AMM family there are several different approaches including constant sum, constant product, curve.finance's approach of a hybrid invariant and balancer's approach of weighted pools. A liquidity provider just deposits a pair of tokens with the exchange's smart contract and receive a proportion of trading fees that happen from traders swapping one token in the pool for the other.

The most successful AMM within the Decentralized Finance sphere is easily Uniswap. Others like Curve have dominated stablecoin pairs but Uniswap was the first true success in the AMM space. With thousands of token pairs and billions of total value locked (TVL) within the Uniswap protocol's smart contracts, it is the largest and most liquid example of a DEX that has seen significant popularity. Uniswap's economic approach is to charge a fixed fee on each swap that occurs with the hope that this is enough to compensate for liquidity providers tying up their funds to act as market makers. These fees vary by pool (stablecoin pools have a fee of 0.05%, "normal" pairs have a fee of 0.3% and exotics have a fee of 1.0%) but not by individual who transacts with these pools.

Swapping tokens through a CPMM can be described as following the $x * y = k$ approach where k is the invariant product of the token balances in the pool before a trade, with x and y being those token balances. The CPMM formula requires that the value of k holds constant pre and post swap (adding/removing liquidity from the pool will change the k value) and thus acts as a pricing curve to determine the quantity of a specified token that will be leaving the pool given the corresponding amount of the other part of the pair that is being added to the pool.

Impermanent Loss

Impermanent loss is defined as the lost opportunity cost for an AMM allowing a swap to happen when compared with if the AMM had just held it's current reserves of the token pair. Fees are required for CPMMs to be appropriately compensated for providing liquidity because impermanent loss is guaranteed under all circumstances without the fees. The unfortunate reality is that for as long as an AMM requires arbitrage to occur to aid in price discovery, some degree of impermanent loss will always happen. The only hope for many pools providing liquidity is that the fees are able to more than compensate for this loss. See [2] for a detailed breakdown of how impermanent loss is calculated for CPMMs under both no fee and with fee scenarios.

Market Microstructure Theory

We borrow heavily from the seminal work by Glosten and Milgrom [1] around defining an economic model that describes the underlying order flow of a securities market as an interaction between informed and uninformed traders. Informed traders are characterized as having perfect knowledge via inside information or arbitrage around where the security's price will next go to. Uninformed traders are defined as having alternative needs for liquidity and/or a long term investment focus and are less concerned about the immediate next movement of the security's price. By definining the observable order flow as a mixture between informed and uninformed trading there is an opportunity to use this framework to better understand the market microstructure of individual liquidity pools in DeFi.

What differentiates this research from the original source material is not just the application of the methodology to cryptocurrency markets but more interestingly that we are using modern statistical algorithms to personalize the market microstructure model down to the transaction level. The original work by Glosten and Milgrom appropriately included concepts like bayesian updating to account for new evidence from the order flow for updating some of the characteristics in the model but it did not contemplate allowing all three components in the model to be fully described down to the swap level. The original research focused on the equities market and defined "swaps" as buys or sells. For our purposes swaps are defined as buys or sells based on the direction from either token **A** to token **B** or vice versa respectively. The implication of this new analysis is that identifying informed trading does not have to happen at the aggregate level across days/weeks/months but down to a much more transactional granularity. The *Generative Process* 0.1 section goes into more detail around the underlying assumptions and mechanics in our statistical model but the main components of the market microstructure model are as follows:

- The prevalence of informed trading is described by the μ parameter
- The bias of uninformed trading $(1 - \mu)$ for being long one token relative to the other is captured via the γ parameter
- The likelihood of a relative increase in token **A**'s value to token **B**'s value is measured via the θ parameter. In our case we define this as the market's implied measurement around this likelihood

All three variables are assumed to be probability measures that are bounded from 0 to 1.

Data and Feature Engineering

0.1 Data

Data for this analysis is pulled via Etherscan's transaction level API with a focus explicitly on several of Uniswap V2's liquidity pools and the swap transactions within them. All transactions through July 2021 on Uniswap were captured and aggregated to a bidirectional swap level. Any transactions which had multiple swaps occurring within them were excluded for these transactions seemed to be bots trading in an unintuitive manner back and forth with themselves.

The list of pairs and some summary statistics that were analyzed are as follows:

Pair	# of Swaps	Mean($\mathbf{A} \to \mathbf{B}$)	Std. Dev($\mathbf{A} \to \mathbf{B}$)	Days of Data	Swaps Per Day
ETH-WBTC	359,951	0.47	0.50	415.04	867.27
ETH-AMPL	353,810	0.46	0.50	433.91	815.39
ETH-LINK	335,982	0.52	0.50	415.00	809.59
ETH-CEL	155,533	0.37	0.48	431.92	360.09
ETH-COMP	147,302	0.55	0.50	403.87	364.73
ETH-1INCH	138,788	0.54	0.50	213.56	649.87
USDC-DAI	134,664	0.51	0.50	419.75	320.82
ETH-REN	125,599	0.48	0.50	433.79	289.54
ETH-MKR	115,646	0.50	0.50	427.01	270.83
FARM-USDC	73,545	0.43	0.50	330.15	222.76
ETH-CHI	72,777	0.54	0.50	430.88	168.90
HEX-USDC	65,094	0.65	0.48	431.88	150.72
WBTC-USDC	61,611	0.50	0.50	431.78	142.69
ETH-PERP	47,325	0.47	0.50	317.20	149.20
ETH-RPL	44,395	0.43	0.49	436.25	101.76
ETH-BOND	36,611	0.47	0.50	256.16	142.92
ETH-POOL	30,536	0.55	0.50	158.77	192.33
ETH-ANT	15,266	0.51	0.50	272.15	56.09
ETH-DGTX	9,356	0.51	0.50	287.17	32.58

0.2 Feature Engineering

The raw data from the Ethereum blockchain contains several useful variables that give specifics around each individual transaction:

- Gas price specified by the trader
- Max gas amount to be paid for each swap
- Transaction time of each swap
- Value of swap in ETH

With these fields we then define additional variables to be used within the analysis that are based upon the order flow of swaps and the speed with which they are happening:

- The number of buys and sells within the last 5 and 10 swaps
- Total value of buys and sells within the last 5 and 10 swaps
- An interaction term between the buy counts of the last 5 and 10 swaps
- The amount of gas paid within the last 5 and 10 swaps

- The amount of trades and sales within the last 1, 5 and 10 minutes for each transaction (excluding any transactions happening on the same block)

The idea behind the inclusion of both the individual transaction level and the order flow type information is to give the model a chance to determine which of these subsets have more predictive power for each of the market microstructure model's three main components.

Markov Chain Monte Carlo Analysis

Markov Chain Monte Carlo (MCMC) methods are an invaluable approach in the bayesian toolbox for infering complex underlying processes that aren't fully observable with data that is available. MCMC methods require the analyst to predefine a system (typically referred to as a generative process) with informed beliefs around the underlying mechanics of how the observed data is generated and the probability distributions of these hidden generators. These methods typically look to minimize the log-likelihood of the generative process by intelligently sampling random numbers to best infer the hidden system and it's different components. A deep dive into the mechanics and theory of MCMC are far beyond the scope of this paper but many excellent overviews exist for the motivated reader [3].

Using the earlier highlighted market microstructure theory's definition of how trades are generated via a mix of informed and uninformed trading we have analyzed the transaction level swap data to better understand the implied dynamics at play within Uniswap's liquidity pools. Informed trading is typically originating from insiders and/or arbitragers while uninformed trading is usually defined as liquidity seekers or longer term investors who aren't especially senstive to small changes in price. By defining our probabilistic framework we are able to determine the following

- The degree of informed trading that occus
- The preference of uninformed trader's to be long one token vs the other in the pair
- The market's implied estimate of the probability of an increase in the value of one token vs the other in the pair

With each of these three components we are able to also discern not just which variables are correlated with high/low levels of each of them but also their relative ranking of importance.

0.1 Generative Process

We define the generative process used within our MCMC model as follows:

The probability of a swap from token **A** to token **B** occurring:

$$\begin{aligned} &P(\boldsymbol{A} \rightarrow \boldsymbol{B}) \sim Bernoulli(y) \\ &y = (\mu + (1-\mu)\gamma)\theta + (1-\mu)\gamma(1-\theta) \end{aligned} \tag{1}$$

The probability that an incoming swap is from an informed trader:

$$\begin{aligned} &\mu = \frac{1}{1 + exp(-\eta_\mu)} \\ &\eta_\mu = \boldsymbol{X_\mu}\beta_\mu \\ &\beta_{\mu,j} \sim \mathcal{N}(0, 1000), j = 1, ..., J_{X_\mu} \\ &\mu \in [0, 1] \end{aligned} \tag{2}$$

The probability that an uninformed trader prefers to swap from token **A** to token **B**:

$$\begin{aligned}
&\gamma = \frac{1}{1 + exp(-\eta_\gamma)} \\
&\eta_\gamma = \boldsymbol{X_\gamma}\beta_\gamma \\
&\beta_{\gamma,j} \sim \mathcal{N}(0, 1000), j = 1, ..., J_{X_\gamma} \\
&\gamma \in [0, 1]
\end{aligned} \tag{3}$$

The probability that token **A**'s value will rise more than token **B**'s:

$$\begin{aligned}
&\theta \sim Beta(1.15, 1.15) \\
&\theta = \frac{1}{1 + exp(-\eta_\theta)} \\
&\eta_\theta = \boldsymbol{X_\theta}\beta_\theta \\
&\beta_{\theta,j} \sim \mathcal{N}(0, 1000), j = 1, ..., J_{X_\theta} \\
&\theta \in [0, 1]
\end{aligned} \tag{4}$$

The log-likelihood function the MCMC algorithm is minimizing:

$$\begin{aligned}
\mathcal{L}(\Theta|\boldsymbol{y}) = p_{Bernoulli}(\boldsymbol{A} \longrightarrow \boldsymbol{B}|y) + \sum_i^{J_{X_\mu}} p(\mathcal{N}(0, 1000)|\beta_{\mu,i}) + \\
\sum_i^{J_{X_\gamma}} p(\mathcal{N}(0, 1000)|\beta_{\gamma,i}) + \sum_i^{J_{X_\theta}} p(\mathcal{N}(0, 1000)|\beta_{\theta,i}) + p(Beta(1.15, 1.15)|\theta)
\end{aligned} \tag{5}$$

From a qualitative perspective the following assumptions are made:

- $P(\boldsymbol{A} \longrightarrow \boldsymbol{B})$'s prior distribution follows a *Bernoulli* distribution [1] as swap transactions can be defined as a 1 for token **A** $\rightarrow$ token **B** and 0 for token **B** $\rightarrow$ token **A** (this can be reversed without loss of generality).
- θ has a prior that follows a *Beta* distribution but this is a weakly informed prior in the sense that we don't have a very strong opinion on the exact shape of this distribution. The 1.15 parameterization ensures that the prior distribution only has a slight "hint" towards the potential shape/mean of θ. The base assumption is built upon the concept that token pair's change in exchange rate follows a random walk via a geometric brownian motion process and thus the probability of an up move vs a down move are roughly equivalent over a short enough time frame.
- μ, γ and θ all can be described by logistic regression models with their own set of covariates.
- β's are normally distributed with a weakly informed prior that is centered at 0 and have a variance of 1000. These beta's are the coefficients within the logistic regression models for each of μ, γ and θ. The three logistic models also include intercept terms.
- $\mathcal{L}(\Theta|\boldsymbol{y})$ defines the log-likelihood function as the sum of the log-likelihood's for the betas of μ, γ and θ against a normal distribution, θ vs it's weakly informed *Beta* prior and the total generative process vs it's *Bernoulli* prior. The prior distributions work to give the MCMC a "hint" about what the underlying process could operate like but if the data provides a strong enough amount of evidence that is on the contrary then we won't be stuck with an unrealistic initial assumption.

0.2 Analysis

The analysis for each token pair used the same approach with the individual data sets being split with the first 75% being the model's training data set and the remaining 25% as our test data set to validate if the model is overfit or not. This split occurs in a sequential manner across time so that the test data set is an out of time hold out. Given the ease with which a complex MCMC algorithm can overfit, having an out of time validation data set ensures that the model's fit truly is robust and not overtraining on too similar of data as can happen with randomly assigned train/test splits.

To give us better insight into the relative importance of each of the three component's variables for making good predictions, the input data sets for the MCMC model are centered and scaled. Both the train and test data sets are centered and scaled on their own to prevent information leakage from our train data set to our test data set. From a statistical perspective centering and scaling our covariates doesn't change the accuracy or robustness of the underlying logistic regression models but it does allow for each model's coefficients to be easily compared as they all have the same mean (due to centering) and the same variance (due to scaling). So a hypothetical variable $\mathbf{X}$ with a coefficient of 0.15 vs a coefficient for variable $\mathbf{Y}$ of 0.075 "roughly" means that variable $\mathbf{X}$ has 2 times higher predictive power than variable $\mathbf{Y}$.

Each token pair's analysis uses the same MCMC algorithm of an Elliptical Slice Sampler [4]. We run the algorithm until the global minimum of the log-likelihood for each converges and we have stable samples with which to imply our underlying generative process. With this model we then score the hold out data set and analyze the consistency of the fit vs the training data set.

Various fit metrics for each token pair's train and test data sets for $P(\boldsymbol{A} \rightarrow \boldsymbol{B})$ were aggregated and it was found that almost all of the models for each pair generalize quite well on the out of time data sets [A]. AUC and F1 scores are very consistent across almost all pairs with only a bit of overfitting happening in the worst case scenario (ETH-CEL for example). Based upon these fit statistics it appears that the generative process we have defined does a reasonably good job of explaining the occurence of different swap directions within each pool.

For the three components (μ, γ and θ) of our generative process we also have the summary metrics for each token pair to highlight the prevalance of informed trading, uninformed trader's bias towards one particular swap direction and the market's implied expectation of likelihood of one token increasing in value relative to the other [B]. Interestingly the presence of informed trading does not appear to be guaranteed within each token pair. Some pairs such as ETH-COMP and ETH-CEL have informed trading happening over 40% of the time while other pairs such as ETH-PERP, ETH-ANT, ETH-RPL and ETH-BOND all have informed trading happening less than 10% of the time. While some relationship appears to exist between volume of swaps and likelihood of informed trading this is not a solid link. If the model is to be believed then some of the more active pools such as ETH-LINK and ETH-WBTC have only a mediocre amount of informed trading happening.

We additionally investigated the three components' coefficients and ranked them across the pairs to understand if there are any consistent trends around what variables have predictive power [C]. One thing to note is that these tables have the intercept for each component removed, signs were removed and the heatmap ranks more to less predictive variables with coloring from green to red respectively. Each token pair's variables within the three model components are ranked within each component. For the informed trading component (μ) of the model it appears that for the vast majority of token pairs analyzed the order flow related variables dominate the importance of the logistic regression models. In a few pairs' size of transaction shows up as having high predictive power but even these pairs still have order flow variables showing up as extremely important. An unintuitive finding is that within the order flow variables the subset focused on buy/trade counts within the prior 5 and 10 minute windows seem to consistently show up as the most predictive variables. Trade counts occurring within the prior 1 or 2 minute windows are still often moderately predictive but on average less so than the longer term windows. For the uninformed trader's bias towards one swap direction or another (γ) the findings from μ not only hold but are even more strong. It appears that the order flow related variables are the only things that matter for this component with the number of buys within the last 5 trades showing up quite frequently as the most important. For the implied likelihood of one token's value rising more than the other (θ) the results are once more consistent with the prior two components. Order flow related variables once again dominate in importance with the buy count interaction and the number of buys within the last 5 and 10 minutes showing up as quite predictive.

More Efficient AMMs

There are numerous opportunities within the DEX space to find more efficient AMMs that more directly address a particular need. It currently appears that the no free lunch paradigm exists where the market as a whole needs informed traders for price discovery but liquidity providers pay a price for getting that price discovery via informed traders making money indirectly off of uninformed traders. This subsidization is well documented in empirical studies and academic research [5], [6]. We hope to move past this common perspective though with a different type of automated market maker that proves to be mutually beneficial for both LPs and uninformed liquidity traders. While informed traders are absolutely needed at a macro level it would be ideal if for an individual trading venue frequented by only LPs and uninformed traders they took their business elsewhere. In this section we propose an extension of one of the more promising new types of AMMs. The proposed modification to this AMM builds on top of the above bayesian analysis and attempts to provide a means to more directly react to changes in order flow.

0.1 DODO EX

Relatively recently the DODO team launched a new AMM dubbed the "Proactive Market Maker" (or PMM from here on out) that leverages an external price oracle and adjustable liquidity parameter to change the curvature of the swap curve [7]. The benefits of this approach is that these two modifications allow for free price discovery by referencing the broader market's price while not incurring the full cost of impermanent loss and a means to more efficiently concentrate liquidity. Despite the inclusion of an external price oracle there are still instances where LPs within DODO are exposed to market risk when large and prolonged price movements occur. The current approach keeps the liquidity parameter, referred to as k, constant across the duration of a pool's lifetime. The subject of our analysis around a more efficient PMM will be focused on how/when to update this liquidity parameter. But first a quick overview of the mechanics of the PMM coming directly from the DODO team.

The PMM is defined as having a price function as follows: $Price = iR$ with $Price$ being the effective swap price, i being the oracle price and R being an offset term that is defined as follows:

1. if $B < B_0, R = 1 - k + (\frac{B_0}{B})^2 k$
2. if $Q < Q_0, R = (1 - k + (\frac{Q_0}{Q})^2)^{-1} k$
3. else $R = 1$

B is defined as the current number of "base" tokens (analagous to our **A** token in the token **A** → token **B** swap) in the pool. Q is defined as the current number of "quote" tokens in the pool. B_0 is the originally deposited number of base tokens supplied by the LPs. And Q_0 is the originally deposited number of quote tokens supplied by LPs. The state of the LP can only ever be in one three aforementioned price definitions.

For a swap to occur the following derivation gives us the quantities to exchange for both $B < B_0$ and $Q < Q_0$:

$$\begin{aligned} \Delta Q &= \int_{B_1}^{B_2} Price\, dB \\ &= \int_{B_1}^{B_2} (1-k)i + i(\frac{B_0}{B})^2\, dB \\ &= i(B_2 - B_1) * (1 - k + k\frac{B_0^2}{B_1 B_2}) \end{aligned} \tag{6}$$

ΔB follows by the same mechanics: $\Delta B = \frac{1}{i}(Q_2 - Q_1) * (1 - k + k\frac{Q_0^2}{Q_1 Q_2})$.

0.2 Adaptive Proactive Market Maker

This work builds upon the insightful analysis from Chitra, Evans and Angeris [9] around the importance of curvature for AMMs and how certain levels of curvature are best suited for different types of markets. Typically stablecoin pools are more suited to low curvature AMMs that allow for less slippage and higher concentration of liquidity around the peg of \$1. Stablecoin pools are characterized as having a lower risk of informed trading as their fair value is usually quite apparent and any small deviation from the peg of \$1 results in arbitragers bringing the price back in line. For regular pairs though there is typically a higher risk of informed trading as pertinent news can take an uncertain amount of time to make it's way out to all members of the market. This risk requires that these types of pairs usually have higher curvature type AMMs.

The findings from this analysis can be generalized to show that there is a direct corollary around how different market conditions are best suited for different curvatures, not just that different coin pair types are best suited by certain curvatures. In quiet markets it very well could be more efficient for a lower curvature non-stablecoin pair to concentrate it's liquidity around the current price. But in more volatile trading regimes it would make sense for the AMM to expect more informed trading and increase it's curvature to compensate for this higher level of risk. This dynamic updating of k will be referred to as the "Adaptive Proactive Market Maker" (or APMM) throughout the rest of this paper and it allows for the flexibility to account for changing market conditions in real time through a straightforward process. Two of the DODO teams' outstanding callouts around areas for them to address are dynamic fees and if k should be changed by the community for liquidity pools. Both of these features are actually addressed with the adoption of something similar to the APMM as changing the liquidity parameter can be viewed as another form of charging a "fee" to the trader [8]. Increasing k results in less coins being swapped all else being equal, similar to increasing the swap fee.

Before we jump into the proposed algorithm we need to define a mathematical model for order flow:

Buy transaction volume is defined as a swap from token **A** to token **B** and denoted as $B(t)$ with t being a predefined time period:

$$\begin{aligned} B(t) &= \sum_{j=1}^{BN(t)} BT_j \\ BN(t) &\sim Poisson(\lambda_B t) \\ BT_j &\sim Exponential(\theta_B) : j \geq 1 \end{aligned} \tag{7}$$

Thus buy transaction volume over t is denoted as $BN(t)$ and is assumed to follow a compound Poisson process where the number of transactions per time period follow an arrival rate of λ_B. This implies that the mean number of transactions over a time period t is equal to $\lambda_B t$. Each buy transaction (BT_i) is assumed to be an independent and identically distributed random variable following an exponential distribution with size θ_B [10]. $BN(t)$ is assumed to be independent of BT. This compound Poisson process has the following properties:

$$\begin{aligned} \mathrm{E}[B(t)] &= \mathrm{E}[\mathrm{E}[B(t)|BN(t)]] \\ &= \mathrm{E}[BN(t)]\mathrm{E}[BT(t)] \\ &= \lambda_B t\theta_B \\ \mathrm{Var}[B(t)] &= \mathrm{E}[\mathrm{Var}[B(t)|BN(t)]] + \mathrm{Var}[\mathrm{E}[B(t)|BN(t)]] \\ &= \mathrm{E}[BN(t)]\mathrm{E}[BT(t)^2] \\ &= 2\lambda_B t\theta_B^2 \end{aligned} \tag{8}$$

Additionally we assumed the same process for sale transactions.

Sell transaction volume is defined as a swap from token **B** to token **A** and denoted as $Q(t)$ with t being a predefined time period:

$$\begin{aligned} Q(t) &= \sum_{j=1}^{QN(t)} QT_j \\ QN(t) &\sim Poisson(\lambda_Q t) \\ QT_j &\sim Exponential(\theta_Q) : j \geq 1 \end{aligned} \tag{9}$$

Sale transaction volume over t is similarly denoted as $QN(t)$ and also is assumed to follow a compound Poisson process where the number of transactions per time period follow an arrival rate of λ_Q. Each sale transaction is assumed to be an independent and identically distributed random variable following an exponential distribution with size $\theta_Q^{'}$. $QN(t)$ is assumed to be independent of QT. $\theta_Q^{'}$ is the rebased size in B's analogous units to allow for a more direct comparison of order flow size. So accounting for our oracle price of i we have $\theta_Q^{'} = \theta_Q/i$. This compound Poisson process has the following properties:

$$
\begin{aligned}
\mathrm{E}[Q(t)] &= \mathrm{E}[\mathrm{E}[Q(t)|QN(t)]] \\
&= \mathrm{E}[QN(t)]\mathrm{E}[QT(t)] \\
&= \lambda_Q t\theta_Q^{'} \\
\mathrm{Var}[Q(t)] &= \mathrm{E}[\mathrm{Var}[Q(t)|QN(t)]] + \mathrm{Var}[\mathrm{E}[Q(t)|QN(t)]] \\
&= \mathrm{E}[QN(t)]\mathrm{E}[QT(t)^2] \\
&= 2\lambda_Q t\theta_Q^{'2}
\end{aligned}
\tag{10}
$$

With our base processes defined we must first note that the compound Poisson processes for both buys and sales actually follow a tweedie distribution which has no closed form formula. Irregardless of this fact we care more about the asymptotic dynamics of the order flow distributions which will each be approximated via the central limit theorem as Normal distributions. Mechanically DODO's smart contract for each pool would only need to track an exponentially decaying average of five parameters: $\lambda_B t$, $\lambda_Q t$, θ_B, $\theta_Q^{'}$ and t the average time period with which these measurements occur. As trades become more spaced out the parameter estimates should converge due to the exponential decay towards a set of realistic equilibrium values. But as the relative speed of trading with the LP picks up, more and more emphasis in the exponential average for these parameters will be placed upon recent trades. We now lay out a very intuitive approach towards updating the liquidity parameter k that leverages the information from the order flow via a classic statistical metric of the similarity between two probability distributions.

Proposition 1 (Dynamic Liquidity Term)**.** *For both $B < B_0$ and $Q > Q_0$ we can define our liquidity term as:*

$$k = min\{1, max\{0, \alpha H(t, \lambda_Q, \lambda_B, \theta_Q^{'}, \theta_B)\}\}$$

with

$$H(t, \lambda_Q, \lambda_B, \theta_Q^{'}, \theta_B) = \sqrt{1 - \sqrt{\frac{2\sqrt{\lambda_Q\lambda_B}\theta_Q^{'}\theta_B}{\lambda_Q\theta_Q^{'2} + \lambda_B\theta_B^2}} e^{-\frac{t}{8}\frac{(\lambda_B\theta_B - \lambda_Q\theta_Q^{'})^2}{\lambda_Q\theta_Q^{'2} + \lambda_B\theta_B^2}}}$$

The risk aversion parameter $\alpha \in \mathbb{R} : \alpha > 0$ is used to scale the sensitivity of the LP to the changes in the order flow. In the proposed modification to DODO's PMM this parameter could be tuned offline and voted on by the DAO to be implemented and periodically updated. Effectively setting $\alpha = 0$ will remove the changes to curvature that the dynamic calculation of k would provide but set $k = 0$. While when $\alpha \to +\infty$ the changes to curvature would stop and the APMM becomes the standard CPMM with $k = 1$. But with α inbetween these two extremes we are able to give more/less emphasis to the dynamic k updating which can help customize to specific LP types (stablecoins vs normal pairs) while still allowing for reaction to changes in order flow.

Heuristic Proof. We use the two order flow distributions to define our liquidity parameter k in an straightforward manner. As the inflows and outflows become more unbalanced, each side of the LP should adapt more aggressively to anticipate further order flows in either directions. Additionally the less uncertainty (i.e. smaller standard deviation) we see for each side of the order flow, the more we should assume there is likely a credible difference in order flow for the LP/broader market. Inflows and outflows in this context are just the usual definitions of **A** $\to$ token **B** and **B** $\to$ token **A** respectively.

In probability theory the Hellinger distance is defined as the difference between two probability distributions f and g [11]. Typically the squared Hellinger distance is written as:

$$
\begin{aligned}
H^2(f,g) &= \frac{1}{2}\int_{-\infty}^{+\infty}(\sqrt{f(x)}-\sqrt{g(x)})^2dx \\
&= \frac{1}{2}\int_{-\infty}^{+\infty} f(x)+g(x)-2\sqrt{f(x)g(x)}dx \\
&= \frac{1}{2}(1+1-2\int_{-\infty}^{+\infty}\sqrt{f(x)g(x)}dx) \\
&= 1-\int_{-\infty}^{+\infty}\sqrt{f(x)g(x)}dx
\end{aligned} \tag{11}
$$

$0 \leq H(f,g) \leq 1$ holds due to first $\int_{-\infty}^{+\infty}(\sqrt{f(x)}-\sqrt{g(x)})^2dx$ being an integral over a squared function which guarantees $H(f,g) \geq 0$. Using the equivalence to $1-\int_{-\infty}^{+\infty}\sqrt{f(x)g(x)}dx$ it holds again that since $f(x)$ and $g(x)$ are both probability density functions they by definition are bounded below by 0. So the integral of $\sqrt{f(x)g(x)}$ is guaranteed to be non negative. At the minimum this then reduces down to the integral of $\sqrt{f(x)g(x)}$ being 0 and thus the upper bound of the Hellinger distance being 1. A Hellinger distance of 1 means that the two distributions are polar opposites of each other while a distance of 0 means that the two distributions are identical.

Using our earlier assumption of $Q(t)$ and $B(t)$ being approximately Normal under the CLT, we get the Hellinger distance formula for two Gaussian distributions of $f(x) \sim \mathcal{N}(\mu_f, \sigma_f^2)$ and $g(x) \sim \mathcal{N}(\mu_g, \sigma_g^2)$:

$$
H^2(f,g) = 1-\sqrt{\frac{2\sigma_f\sigma_g}{\sigma_f^2+\sigma_g^2}}e^{-\frac{1}{4}\frac{(\mu_f-\mu_g)^2}{\sigma_f^2+\sigma_g^2}} \tag{12}
$$

Substituting in the mean and variance of both the compound Poisson processes $Q(t)$ and $B(t)$ into the Hellinger distance formula and making some minor simplifications we arrive at [1] listed above. □

Corollary 1.1. *If we are instead looking for a simpler version of the order flow model then the size of the inflow/outflow transactions can dropped from the compound Poisson processes which just reduces to plain Poisson processes. This leaves us with only $\lambda_B t$ and $\lambda_Q t$ in our model. A less computationally expensive version of the Hellinger distance formula exists if the two probability distributions are Poisson $H^2(f,g) = 1-e^{-\frac{1}{2}(\sqrt{\lambda_f}-\sqrt{\lambda_f})^2}$. Replacing the λ's from the Poisson processes into the Hellinger distance formula now gives us:*

$$
H(t,\lambda_B,\lambda_Q) = \sqrt{1-e^{-\frac{t}{2}(\sqrt{\lambda_B}-\sqrt{\lambda_Q})^2}}
$$

This more simplistic model is more closely aligned to the gist of the earlier MCMC analysis in that it only focuses on the directional component of order flow and ignores the size of the transactions.

Now armed with a closed form solution for answering our question around how to quantify the difference in order flow we can make several important observations. Uninformed trading should not be especially directional and we would expect there to be low autocorrelation between the direction/size of one swap to the next. This is effectively the very definition of uninformed trading. Within the Hellinger distance formula this is fully represented in that if swaps in both directions have roughly the same arrival rate of swaps and same size of trades then our k parameter (assuming $\alpha = 1$) will approach 0. On the other hand if informed trading is occurring then we'd expect there to be a difference between the swap directions' arrival rate and/or transaction size. The Hellinger distance will only equal 0 when both distributions are identical so we effectively have both the difference in the means and standard deviations of the two distributions to fully reflect the degree of difference that exists. A conceptual example of this would be that if there is a wide spread liquidation from centralized exchanges making margin calls then a flight to safety could occur and swaps would likely be in just one direction of being towards to the more liquid/stable of a pair such as ETH to DAI. This change in order flow would rapidly be picked up by the proposed algorithm which would virtually remove concentration of a given LP's liquidity around the current price through it's k parameter approaching 1.

One subtle component of the APMM is that the t parameter of the poisson arrival processes does influence the curvature in a very important way. A larger t means that we are supplying liquidity over a longer expected time period and that will require a higher curvature in the AMM. Even if all arrival rates ($\lambda's$) and transaction sizes ($\theta's$) variables are held constant, expecting the time between two swaps to increase through a larger relative t will increase the APMM's uncertainty around expected order flow. Intuitively we would want the curvature to automatically increase to compensate for this increased informational risk and this algorithm achieves exactly that.

Conclusion

Within this analysis we have identified the prevalance of informed and uninformed trading occurring in cryptocurrency AMMs through Markov Chain Monte Carlo methods. The key findings were that highly predictive statistical models could be built with this approach and across all LP pairs the order flow related characteristics had high statistical importance. With these new insights we have proposed a new AMM design that builds on top of the DODO team's Proactive Market Maker framework. Through the tracking of meta data around swap direction arrival rates and swap sizes as well as the application of a well known measure of statistical distance it would be technically feasible for a smart contract to approximate how many professional market makers manage their inventory in markets dominated by limit order books. We hope that even if the implementation of the APMM does not pan for the DODO team then it at least provides a new perspective on the possibility of including statistical learning methodologies into smart contracts throughout the vibrant DeFi ecosystem. Additionally, this approach can potentially prove to be mutually beneficial for market participants with uninformed liquidity traders getting better prices on AMMs while allowing liquidity providers to achieve fair returns without excessive market or impermanent loss risk.

Some additional areas of research related to this work that could be explored in more detail:

- These type of Markov Chain Monte Carlo models can be used in real time to identify when informed order flows are likely occuring on different DEXes. These orders can then be piggy backed before the informed transactions have been arbitraged away across other markets.
- Simplify the APMM to be more gas efficient while still maintaining the ethos of what it is trying to achieve. Likely there are opportunities to dramatically simplify the full Hellinger formula under certain scenarios. Even more parsimonious options likely exist such as a regime characterizing set of rules that could be coded as simple if-then logic.
- Create a better AMM that more naturally integrates information around order flow instead of the work around with the dynamic k variable.
- Expand upon the work from the DODO team and Chitra, Evans and Angeris to make curvature more pronounced under different market conditions. The CPMM may not be the most extreme that an AMM should be converging towards. Additionally it may make sense to have the liquidity parameter k vary depending on which direction the swapping is happening.

Appendices

A MCMC Model Fit

AUC and F1 scores by token pair

Pair	Train F1	Train AUC	Test F1	Test AUC
ETH-LINK	0.75	0.85	0.75	0.84
USDC-DAI	0.75	0.86	0.72	0.82
ETH-WBTC	0.73	0.84	0.72	0.83
ETH-MKR	0.72	0.83	0.74	0.83
ETH-COMP	0.81	0.85	0.77	0.81
ETH-CEL	0.72	0.88	0.77	0.83
HEX-USDC	0.84	0.83	0.87	0.83
ETH-POOL	0.76	0.81	0.84	0.77
ETH-1INCH	0.78	0.80	0.72	0.77
ETH-AMPL	0.69	0.83	0.74	0.79
ETH-PERP	0.63	0.79	0.68	0.78
ETH-REN	0.71	0.84	0.75	0.82
ETH-ANT	0.73	0.79	0.75	0.78
FARM-USDC	0.70	0.87	0.73	0.80
ETH-DGTX	0.73	0.82	0.75	0.80
ETH-RPL	0.64	0.80	0.58	0.80
ETH-BOND	0.68	0.80	0.69	0.79
ETH-CHI	0.75	0.83	0.84	0.83

B MCMC Model Component Statistics

Summary statistics for μ, γ and θ

Pair	mean(μ)	sd(μ)	mean(γ)	sd(γ)	mean(θ)	sd(θ)
ETH-LINK	0.18	0.20	0.46	0.33	0.55	0.07
USDC-DAI	0.23	0.25	0.46	0.36	0.53	0.06
ETH-WBTC	0.16	0.15	0.50	0.33	0.48	0.05
ETH-MKR	0.29	0.23	0.42	0.33	0.56	0.07
ETH-COMP	0.42	0.30	0.70	0.33	0.42	0.10
ETH-CEL	0.46	0.40	0.22	0.29	0.55	0.12
HEX-USDC	0.19	0.17	0.63	0.31	0.49	0.04
ETH-POOL	0.12	0.09	0.54	0.29	0.51	0.03
ETH-1INCH	0.11	0.19	0.57	0.27	0.49	0.05
ETH-AMPL	0.12	0.26	0.45	0.30	0.50	0.08
ETH-PERP	0.07	0.12	0.46	0.26	0.50	0.03
ETH-REN	0.32	0.26	0.37	0.33	0.57	0.09
ETH-ANT	0.07	0.08	0.50	0.27	0.50	0.03
FARM-USDC	0.14	0.27	0.36	0.30	0.53	0.08
ETH-DGTX	0.13	0.06	0.52	0.29	0.51	0.05
ETH-RPL	0.07	0.14	0.44	0.27	0.50	0.04
ETH-BOND	0.07	0.14	0.48	0.26	0.49	0.04
ETH-CHI	0.17	0.14	0.54	0.33	0.48	0.04

C MCMC Model Coefficient Information

Breakdown of logistic regression models' centered and scaled coefficients

Model Coefficient Heatmap for Mu Pt. 1

Names	ETH-LINK	USDC-DAI	ETH-WBTC	ETH-MKR	ETH-COMP	ETH-CHI
gasPrice	0.24	0.26	0.21	0.31	0.39	0.21
maxFee	0.15	0.29	0.28	0.14	0.47	1.40
gasLimit	0.04	0.23	0.13	0.05	0.23	0.12
sizeOfTransaction	0.58	0.63	0.02	2.63	0.13	1.56
secondsSinceLastSwap	0.00	0.05	0.07	0.03	0.06	0.08
buyCountLast5Trades	3.57	4.19	1.78	4.09	0.17	0.66
buyCountLast10Trades	0.47	0.73	1.82	0.96	2.75	2.54
buyCountInteraction	3.70	4.68	3.06	4.83	5.11	2.50
buyValueLast5Trades	0.69	0.06	0.77	0.32	0.09	0.68
buyValueLast10Trades	0.72	0.17	0.35	0.22	0.10	0.10
sellValueLast5Trades	0.12	0.55	0.58	0.69	0.08	0.41
sellValueLast10Trades	0.47	0.18	0.04	0.06	0.10	0.24
sumOfGasLast5Trades	0.00	0.03	0.04	0.08	0.01	0.07
sumOfGasLast10Trades	0.18	0.19	0.10	0.01	0.24	0.20
tradesWithinLast1Minutes	0.11	0.93	1.93	0.15	1.10	0.79
buysWithinLast1Minutes	0.24	0.53	0.25	0.68	0.92	0.01
tradesWithinLast2Minutes	0.01	1.28	2.34	0.13	3.27	1.31
buysWithinLast2Minutes	1.99	0.26	0.22	1.26	2.08	0.59
tradesWithinLast5Minutes	1.32	1.14	1.67	1.04	6.36	1.53
buysWithinLast5Minutes	4.79	2.46	0.31	3.00	4.49	0.16
tradesWithinLast10Minutes	2.21	0.71	0.65	1.30	6.25	0.98
buysWithinLast10Minutes	4.49	4.15	1.30	4.43	4.47	0.46

Model Coefficient Heatmap for Mu Pt. 2

Names	ETH-COMP	ETH-CEL	HEX-USDC	ETH-POOL	ETH-1INCH	ETH-CHI
gasPrice	0.39	0.74	0.25	0.06	0.25	0.21
maxFee	0.47	0.07	0.16	0.25	0.27	1.40
gasLimit	0.23	0.05	0.44	0.14	0.17	0.12
sizeOfTransaction	0.13	1.62	0.94	0.81	2.08	1.56
secondsSinceLastSwap	0.06	0.14	0.12	0.31	0.23	0.08
buyCountLast5Trades	0.17	8.21	0.07	0.14	0.15	0.66
buyCountLast10Trades	2.75	1.59	1.31	0.68	0.37	2.54
buyCountInteraction	5.11	0.50	2.59	0.63	2.00	2.50
buyValueLast5Trades	0.09	0.78	0.76	0.40	0.50	0.68
buyValueLast10Trades	0.10	0.13	0.02	0.04	0.10	0.10
sellValueLast5Trades	0.08	0.28	1.75	0.64	0.59	0.41
sellValueLast10Trades	0.10	0.01	0.82	0.12	0.08	0.24
sumOfGasLast5Trades	0.01	0.53	0.36	0.11	0.03	0.07
sumOfGasLast10Trades	0.24	0.49	0.01	0.16	0.01	0.20
tradesWithinLast1Minutes	1.10	0.36	0.29	0.29	1.02	0.79
buysWithinLast1Minutes	0.92	0.51	0.26	0.22	0.23	0.01
tradesWithinLast2Minutes	3.27	0.39	0.45	0.61	1.16	1.31
buysWithinLast2Minutes	2.08	2.06	0.37	0.62	0.88	0.59
tradesWithinLast5Minutes	6.36	2.21	1.70	0.62	0.66	1.53
buysWithinLast5Minutes	4.49	7.12	0.23	0.66	0.62	0.16
tradesWithinLast10Minutes	6.25	2.81	2.36	0.38	0.85	0.98
buysWithinLast10Minutes	4.47	6.68	0.50	0.05	0.55	0.46

Model Coefficient Heatmap for Mu Pt. 3

Names	ETH-1INCH	ETH-AMPL	ETH-PERP	ETH-REN	ETH-ANT	ETH-CHI
gasPrice	0.25	0.16	0.36	0.00	0.17	0.21
maxFee	0.27	0.04	0.75	0.01	0.22	1.40
gasLimit	0.17	0.18	0.28	0.08	0.13	0.12
sizeOfTransaction	2.08	2.31	0.23	1.96	0.02	1.56
secondsSinceLastSwap	0.23	0.51	0.13	0.11	0.27	0.08
buyCountLast5Trades	0.15	0.30	0.91	4.83	0.29	0.66
buyCountLast10Trades	0.37	1.52	0.89	1.18	0.24	2.54
buyCountInteraction	2.00	0.39	0.62	4.53	0.06	2.50
buyValueLast5Trades	0.50	0.23	0.45	0.33	0.01	0.68
buyValueLast10Trades	0.10	0.09	0.03	0.00	0.06	0.10
sellValueLast5Trades	0.59	0.25	0.31	0.24	0.62	0.41
sellValueLast10Trades	0.08	0.03	0.02	0.28	0.02	0.24
sumOfGasLast5Trades	0.03	0.43	0.46	0.22	0.15	0.07
sumOfGasLast10Trades	0.01	0.07	0.11	0.14	0.29	0.20
tradesWithinLast1Minutes	1.02	3.49	0.18	0.19	0.06	0.79
buysWithinLast1Minutes	0.23	0.63	0.28	0.29	0.39	0.01
tradesWithinLast2Minutes	1.16	2.16	0.09	0.40	0.14	1.31
buysWithinLast2Minutes	0.88	0.31	0.05	0.75	0.37	0.59
tradesWithinLast5Minutes	0.66	1.20	0.30	1.43	0.46	1.53
buysWithinLast5Minutes	0.62	0.21	0.72	3.19	0.19	0.16
tradesWithinLast10Minutes	0.85	0.93	0.48	3.22	0.47	0.98
buysWithinLast10Minutes	0.55	0.17	1.31	6.19	0.18	0.46

Model Coefficient Heatmap for Mu Pt. 4

Names	ETH-ANT	FARM-USDC	ETH-DGTX	ETH-RPL	ETH-BOND	ETH-CHI
gasPrice	0.17	0.05	0.20	0.32	0.15	0.21
maxFee	0.22	0.03	0.29	0.13	0.12	1.40
gasLimit	0.13	0.98	0.33	0.34	0.30	0.12
sizeOfTransaction	0.02	0.81	0.45	0.30	0.96	1.56
secondsSinceLastSwap	0.27	0.30	0.07	0.30	0.05	0.08
buyCountLast5Trades	0.29	0.89	0.03	0.07	0.19	0.66
buyCountLast10Trades	0.24	1.56	0.57	0.36	0.05	2.54
buyCountInteraction	0.06	1.30	0.47	0.80	0.22	2.50
buyValueLast5Trades	0.01	0.35	0.63	0.43	1.03	0.68
buyValueLast10Trades	0.06	0.01	0.01	0.87	0.51	0.10
sellValueLast5Trades	0.62	0.34	0.38	0.54	0.60	0.41
sellValueLast10Trades	0.02	0.03	0.03	0.49	0.13	0.24
sumOfGasLast5Trades	0.15	0.17	0.53	0.25	0.15	0.07
sumOfGasLast10Trades	0.29	0.24	0.25	0.11	0.29	0.20
tradesWithinLast1Minutes	0.06	1.27	0.24	0.63	0.91	0.79
buysWithinLast1Minutes	0.39	1.18	0.14	0.19	0.44	0.01
tradesWithinLast2Minutes	0.14	1.44	0.23	1.07	0.69	1.31
buysWithinLast2Minutes	0.37	1.51	0.13	0.12	0.38	0.59
tradesWithinLast5Minutes	0.46	1.53	0.41	0.78	0.93	1.53
buysWithinLast5Minutes	0.19	0.37	0.28	0.03	0.00	0.16
tradesWithinLast10Minutes	0.47	1.25	0.06	0.89	0.27	0.98
buysWithinLast10Minutes	0.18	1.33	0.02	0.56	0.05	0.46

Model Coefficient Heatmap for Gamma Pt. 1

Names	ETH-LINK	USDC-DAI	ETH-WBTC	ETH-MKR	ETH-COMP	ETH-CHI
gasPrice	0.74	0.16	0.02	0.15	0.00	0.59
maxFee	0.41	0.24	0.31	0.19	0.03	0.38
gasLimit	0.19	0.10	0.54	0.68	0.15	0.02
sizeOfTransaction	0.51	0.16	1.42	0.28	0.56	0.44
secondsSinceLastSwap	0.04	0.01	0.08	0.12	0.03	0.09
buyCountLast5Trades	4.81	5.65	4.27	4.91	3.33	2.81
buyCountLast10Trades	2.39	1.27	0.67	1.37	2.00	0.13
buyCountInteraction	0.98	0.40	0.41	0.81	2.05	2.61
buyValueLast5Trades	0.06	0.05	0.21	0.14	0.19	0.28
buyValueLast10Trades	0.01	0.02	0.16	0.10	0.10	0.10
sellValueLast5Trades	0.15	0.08	0.08	0.28	0.02	0.14
sellValueLast10Trades	0.20	0.04	0.07	0.14	0.02	0.13
sumOfGasLast5Trades	0.06	0.10	0.05	0.01	0.03	0.31
sumOfGasLast10Trades	0.19	0.05	0.05	0.04	0.14	0.28
tradesWithinLast1Minutes	0.87	0.62	0.64	0.66	1.28	1.28
buysWithinLast1Minutes	1.02	0.79	0.88	0.65	1.87	1.35
tradesWithinLast2Minutes	1.82	0.57	1.25	0.99	1.85	1.30
buysWithinLast2Minutes	1.90	0.58	1.63	1.02	2.44	1.50
tradesWithinLast5Minutes	2.78	2.03	1.18	1.11	1.30	1.34
buysWithinLast5Minutes	2.87	2.18	1.44	1.25	1.79	1.27
tradesWithinLast10Minutes	1.77	1.96	0.69	1.83	0.36	0.13
buysWithinLast10Minutes	1.68	2.31	0.52	1.69	0.50	0.95

Model Coefficient Heatmap for Gamma Pt. 2

Names	ETH-COMP	ETH-CEL	HEX-USDC	ETH-POOL	ETH-1INCH	ETH-CHI
gasPrice	0.00	0.25	0.04	0.97	0.20	0.59
maxFee	0.03	0.21	0.00	0.53	0.22	0.38
gasLimit	0.15	0.12	0.07	0.29	0.03	0.02
sizeOfTransaction	0.56	0.32	0.92	0.38	0.68	0.44
secondsSinceLastSwap	0.03	0.35	0.05	0.12	0.04	0.09
buyCountLast5Trades	3.33	4.32	3.46	2.34	2.15	2.81
buyCountLast10Trades	2.00	3.24	0.07	0.47	0.06	0.13
buyCountInteraction	2.05	0.73	1.30	1.06	1.22	2.61
buyValueLast5Trades	0.19	0.16	0.40	0.20	0.14	0.28
buyValueLast10Trades	0.10	0.07	0.11	0.09	0.00	0.10
sellValueLast5Trades	0.02	0.03	0.44	0.18	0.09	0.14
sellValueLast10Trades	0.02	0.17	0.02	0.05	0.00	0.13
sumOfGasLast5Trades	0.03	0.05	0.01	0.35	0.00	0.31
sumOfGasLast10Trades	0.14	0.07	0.00	0.00	0.02	0.28
tradesWithinLast1Minutes	1.28	0.87	0.05	1.34	4.21	1.28
buysWithinLast1Minutes	1.87	0.98	0.05	1.56	3.73	1.35
tradesWithinLast2Minutes	1.85	1.73	0.04	1.68	2.75	1.30
buysWithinLast2Minutes	2.44	1.54	0.13	1.68	2.57	1.50
tradesWithinLast5Minutes	1.30	1.65	1.30	1.15	1.01	1.34
buysWithinLast5Minutes	1.79	1.61	1.46	1.46	1.31	1.27
tradesWithinLast10Minutes	0.36	0.41	2.40	0.95	0.64	0.13
buysWithinLast10Minutes	0.50	0.01	2.69	0.65	1.17	0.95

Model Coefficient Heatmap for Gamma Pt. 3

Names	ETH-1INCH	ETH-AMPL	ETH-PERP	ETH-REN	ETH-ANT	ETH-CHI
gasPrice	0.20	0.61	0.20	0.04	0.31	0.59
maxFee	0.22	0.10	0.51	0.14	0.06	0.38
gasLimit	0.03	0.05	0.37	0.23	0.02	0.02
sizeOfTransaction	0.68	0.20	0.12	0.10	1.05	0.44
secondsSinceLastSwap	0.04	0.12	0.02	0.26	0.07	0.09
buyCountLast5Trades	2.15	4.64	2.78	5.18	2.19	2.81
buyCountLast10Trades	0.06	1.69	0.22	2.50	0.01	0.13
buyCountInteraction	1.22	1.59	0.22	1.21	0.79	2.61
buyValueLast5Trades	0.14	0.28	0.09	0.12	0.19	0.28
buyValueLast10Trades	0.00	0.05	0.09	0.08	0.21	0.10
sellValueLast5Trades	0.09	0.25	0.00	0.28	0.34	0.14
sellValueLast10Trades	0.00	0.06	0.01	0.21	0.11	0.13
sumOfGasLast5Trades	0.00	0.42	0.12	0.01	0.09	0.31
sumOfGasLast10Trades	0.02	0.02	0.12	0.05	0.12	0.28
tradesWithinLast1Minutes	4.21	2.82	1.39	0.88	1.01	1.28
buysWithinLast1Minutes	3.73	3.04	1.32	0.90	1.40	1.35
tradesWithinLast2Minutes	2.75	2.92	1.54	1.39	0.70	1.30
buysWithinLast2Minutes	2.57	2.96	1.35	1.33	0.86	1.50
tradesWithinLast5Minutes	1.01	1.35	0.99	2.48	1.06	1.34
buysWithinLast5Minutes	1.31	1.37	1.54	2.31	1.10	1.27
tradesWithinLast10Minutes	0.64	0.72	1.07	1.91	1.22	0.13
buysWithinLast10Minutes	1.17	0.66	1.60	1.82	1.27	0.95

Model Coefficient Heatmap for Gamma Pt. 4

Names	ETH-ANT	FARM-USDC	ETH-DGTX	ETH-RPL	ETH-BOND	ETH-CHI
gasPrice	0.31	0.27	0.24	0.30	0.70	0.59
maxFee	0.06	0.58	0.49	0.27	0.41	0.38
gasLimit	0.02	2.48	0.58	0.02	0.07	0.02
sizeOfTransaction	1.05	0.42	0.64	0.84	0.59	0.44
secondsSinceLastSwap	0.07	0.02	0.00	0.11	0.01	0.09
buyCountLast5Trades	2.19	4.49	2.53	3.06	2.67	2.81
buyCountLast10Trades	0.01	1.19	0.09	0.62	0.39	0.13
buyCountInteraction	0.79	0.72	1.12	0.51	0.08	2.61
buyValueLast5Trades	0.19	0.17	0.13	0.07	0.30	0.28
buyValueLast10Trades	0.21	0.07	0.06	0.06	0.05	0.10
sellValueLast5Trades	0.34	0.17	0.08	0.05	0.10	0.14
sellValueLast10Trades	0.11	0.08	0.06	0.13	0.04	0.13
sumOfGasLast5Trades	0.09	0.12	0.19	0.08	0.25	0.31
sumOfGasLast10Trades	0.12	0.44	0.20	0.02	0.02	0.28
tradesWithinLast1Minutes	1.01	0.67	0.16	1.50	1.78	1.28
buysWithinLast1Minutes	1.40	0.80	0.24	1.26	1.77	1.35
tradesWithinLast2Minutes	0.70	1.56	0.00	1.18	1.78	1.30
buysWithinLast2Minutes	0.86	2.97	0.36	0.87	1.81	1.50
tradesWithinLast5Minutes	1.06	3.66	0.71	1.50	2.06	1.34
buysWithinLast5Minutes	1.10	6.61	0.55	1.00	1.90	1.27
tradesWithinLast10Minutes	1.22	3.69	0.59	1.74	1.70	0.13
buysWithinLast10Minutes	1.27	5.57	0.74	2.45	1.65	0.95

Model Coefficient Heatmap for Theta Pt. 1

Names	ETH-LINK	USDC-DAI	ETH-WBTC	ETH-MKR	ETH-COMP	ETH-CHI
gasPrice	0.05	0.03	0.06	0.01	0.14	0.14
maxFee	0.03	0.01	0.04	0.02	0.10	0.11
gasLimit	0.07	0.02	0.03	0.05	0.07	0.03
sizeOfTransaction	0.03	0.16	0.03	0.19	0.01	0.13
secondsSinceLastSwap	0.01	0.02	0.01	0.03	0.05	0.04
buyCountLast5Trades	1.20	0.90	0.08	1.23	0.63	0.02
buyCountLast10Trades	0.50	0.06	0.23	0.23	0.23	0.36
buyCountInteraction	1.28	0.55	0.42	0.99	0.42	0.49
buyValueLast5Trades	0.01	0.04	0.05	0.00	0.02	0.06
buyValueLast10Trades	0.01	0.00	0.03	0.01	0.00	0.04
sellValueLast5Trades	0.06	0.03	0.03	0.08	0.04	0.12
sellValueLast10Trades	0.00	0.02	0.00	0.03	0.06	0.00
sumOfGasLast5Trades	0.01	0.00	0.01	0.00	0.06	0.01
sumOfGasLast10Trades	0.01	0.04	0.02	0.02	0.09	0.03
tradesWithinLast1Minutes	0.02	0.06	0.08	0.18	0.21	0.23
buysWithinLast1Minutes	0.00	0.05	0.08	0.17	0.07	0.20
tradesWithinLast2Minutes	0.02	0.04	0.14	0.08	0.06	0.17
buysWithinLast2Minutes	0.07	0.03	0.06	0.03	0.20	0.20
tradesWithinLast5Minutes	0.04	0.16	0.38	0.02	0.39	0.42
buysWithinLast5Minutes	0.31	0.10	0.45	0.16	0.35	0.30
tradesWithinLast10Minutes	0.82	0.63	0.19	1.07	1.77	0.31
buysWithinLast10Minutes	0.83	0.71	0.24	1.12	1.60	0.47

Model Coefficient Heatmap for Theta Pt. 2

Names	ETH-COMP	ETH-CEL	HEX-USDC	ETH-POOL	ETH-1INCH	ETH-CHI
gasPrice	0.14	0.10	0.08	0.01	0.03	0.14
maxFee	0.10	0.07	0.05	0.05	0.04	0.11
gasLimit	0.07	0.05	0.03	0.02	0.04	0.03
sizeOfTransaction	0.01	0.12	0.11	0.06	0.17	0.13
secondsSinceLastSwap	0.05	0.11	0.00	0.06	0.01	0.04
buyCountLast5Trades	0.63	0.84	0.32	0.04	0.68	0.02
buyCountLast10Trades	0.23	0.17	0.24	0.14	0.35	0.36
buyCountInteraction	0.42	0.48	0.06	0.14	0.75	0.49
buyValueLast5Trades	0.02	0.00	0.11	0.03	0.06	0.06
buyValueLast10Trades	0.00	0.04	0.08	0.02	0.08	0.04
sellValueLast5Trades	0.04	0.05	0.06	0.04	0.09	0.12
sellValueLast10Trades	0.06	0.11	0.02	0.01	0.01	0.00
sumOfGasLast5Trades	0.06	0.03	0.05	0.11	0.05	0.01
sumOfGasLast10Trades	0.09	0.01	0.02	0.03	0.06	0.03
tradesWithinLast1Minutes	0.21	0.09	0.06	0.02	0.04	0.23
buysWithinLast1Minutes	0.07	0.16	0.08	0.12	0.03	0.20
tradesWithinLast2Minutes	0.06	0.03	0.05	0.31	0.09	0.17
buysWithinLast2Minutes	0.20	0.05	0.05	0.20	0.10	0.20
tradesWithinLast5Minutes	0.39	0.67	0.18	0.33	0.02	0.42
buysWithinLast5Minutes	0.35	0.41	0.17	0.12	0.12	0.30
tradesWithinLast10Minutes	1.77	1.62	0.65	0.13	0.12	0.31
buysWithinLast10Minutes	1.60	1.84	0.69	0.27	0.21	0.47

Model Coefficient Heatmap for Theta Pt. 3

Names	ETH-1INCH	ETH-AMPL	ETH-PERP	ETH-REN	ETH-ANT	ETH-CHI
gasPrice	0.03	0.00	0.04	0.11	0.01	0.14
maxFee	0.04	0.01	0.02	0.06	0.07	0.11
gasLimit	0.04	0.00	0.05	0.06	0.02	0.03
sizeOfTransaction	0.17	0.11	0.17	0.13	0.01	0.13
secondsSinceLastSwap	0.01	0.00	0.05	0.06	0.03	0.04
buyCountLast5Trades	0.68	0.50	0.03	1.61	0.26	0.02
buyCountLast10Trades	0.35	0.13	0.17	0.45	0.11	0.36
buyCountInteraction	0.75	0.13	0.25	1.45	0.25	0.49
buyValueLast5Trades	0.06	0.08	0.00	0.01	0.06	0.06
buyValueLast10Trades	0.08	0.04	0.00	0.01	0.13	0.04
sellValueLast5Trades	0.09	0.25	0.09	0.04	0.00	0.12
sellValueLast10Trades	0.01	0.11	0.04	0.06	0.09	0.00
sumOfGasLast5Trades	0.05	0.01	0.07	0.04	0.07	0.01
sumOfGasLast10Trades	0.06	0.01	0.00	0.02	0.01	0.03
tradesWithinLast1Minutes	0.04	0.05	0.05	0.15	0.01	0.23
buysWithinLast1Minutes	0.03	0.02	0.01	0.23	0.18	0.20
tradesWithinLast2Minutes	0.09	0.11	0.10	0.07	0.24	0.17
buysWithinLast2Minutes	0.10	0.10	0.16	0.00	0.39	0.20
tradesWithinLast5Minutes	0.02	0.10	0.30	0.08	0.29	0.42
buysWithinLast5Minutes	0.12	0.10	0.46	0.09	0.24	0.30
tradesWithinLast10Minutes	0.12	0.25	0.13	0.77	0.24	0.31
buysWithinLast10Minutes	0.21	0.17	0.10	0.93	0.16	0.47

Model Coefficient Heatmap for Theta Pt. 4

Names	ETH-ANT	FARM-USDC	ETH-DGTX	ETH-RPL	ETH-BOND	ETH-CHI
gasPrice	0.01	0.03	0.11	0.07	0.00	0.14
maxFee	0.07	0.04	0.03	0.08	0.06	0.11
gasLimit	0.02	0.01	0.04	0.02	0.00	0.03
sizeOfTransaction	0.01	0.17	0.13	0.11	0.22	0.13
secondsSinceLastSwap	0.03	0.03	0.03	0.01	0.02	0.04
buyCountLast5Trades	0.26	0.26	0.31	0.39	0.12	0.02
buyCountLast10Trades	0.11	0.43	0.13	0.23	0.05	0.36
buyCountInteraction	0.25	1.35	0.11	0.39	0.07	0.49
buyValueLast5Trades	0.06	0.05	0.15	0.11	0.15	0.06
buyValueLast10Trades	0.13	0.05	0.01	0.04	0.06	0.04
sellValueLast5Trades	0.00	0.01	0.22	0.02	0.07	0.12
sellValueLast10Trades	0.09	0.02	0.37	0.03	0.12	0.00
sumOfGasLast5Trades	0.07	0.04	0.04	0.08	0.05	0.01
sumOfGasLast10Trades	0.01	0.08	0.07	0.00	0.03	0.03
tradesWithinLast1Minutes	0.01	0.33	0.19	0.12	0.16	0.23
buysWithinLast1Minutes	0.18	0.35	0.15	0.06	0.14	0.20
tradesWithinLast2Minutes	0.24	0.12	0.10	0.03	0.05	0.17
buysWithinLast2Minutes	0.39	0.19	0.31	0.06	0.07	0.20
tradesWithinLast5Minutes	0.29	0.09	0.24	0.14	0.01	0.42
buysWithinLast5Minutes	0.24	0.00	0.29	0.23	0.13	0.30
tradesWithinLast10Minutes	0.24	0.38	0.30	0.41	0.27	0.31
buysWithinLast10Minutes	0.16	0.30	0.25	0.15	0.42	0.47

References

[1] Lawrence R. Glosten and Paul R. Milgrom, "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders", *Journal of Financial Economics*, Volume 14, Issue 1, 1985, Pages 71-100

[2] Tian, Ryan. "Understanding the XYK Model of Pooled Liquidity." *Medium (blog)*, January 1, 2020. https://medium.com/phoenix-finance/understanding-the-xyk-model-of-pooled-liquidity-7340fdc20d9c.

[3] Gamerman, Dani, and Hedibert Freitas Lopes. 2006. Markov chain Monte Carlo: stochastic simulation for Bayesian inference. Boca Raton: Taylor Francis.

[4] Murray, I., Adams, R. MacKay, D.. (2010). "Elliptical slice sampling. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics", *Proceedings of Machine Learning Research* 9:541-548 Available from http://proceedings.mlr.press/v9/murray10a.html .

[5] OHara, M. (1995) Market Microstructure Theory. Blackwell, Cambridge.

[6] Foucault, Thierry, Pagano, Marco and Roell, Ailsa, (2013), Market Liquidity: Theory, Evidence, and Policy, Oxford University Press.

[7] DODO Team. "DODO Whitepaper" *GitHub* https://dodoex.github.io/docs/docs/whitepaper/

[8] Radar Bear. "A Gentle Introduction to How DODO Deals with Market Risk." *Medium (blog)*, August 25, 2020. https://medium.com/dodoex/a-gentle-introduction-to-how-dodo-deals-with-market-risk-39dac80f740b

[9] Angeris, Guillermo, Alex Evans, and Tarun Chitra. "When does the tail wag the dog? Curvature and market making." arXiv preprint arXiv:2012.08040 (2020).

[10] Wikipedia contributors, "Compound Poisson process," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/none?title=Compound_Poisson_processoldid=964442685 (accessed September 1, 2021).

[11] Wikipedia contributors, "Hellinger distance," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/none?title=Hellinger_distanceoldid=1001077827 (accessed September 1, 2021).