Discrimination of Toxic Flow in Uniswap V3: Part 1

CrocSwap

This post is a new installment in an ongoing series by @0xfbifemboy on Uniswap liquidity pools, concentrated liquidity, and fee dynamics. It is the first of multiple posts in a subsequence which aims to focus on the characterization of toxic flow in ETH/USDC swap data and potential implementations of price discrimination or flow segmentation mechanisms.

Introduction

In a prior research post, we explored the profitability of Uniswap ETH/USDC liquidity providers by calculating markouts relative to future Binance prices. As part of this analysis, we obtain a measure of PnL on a per-swap basis, which allowed us to determine the degree to which swap profitability varied according to the notional trade size.

In this post, we look more closely at the specific predictors of swap PnL. We are motivated by the notion of price discrimination, where an enterprising DEX might be able to "win" trade flow by charging below-market rates to non-toxic flow or, conversely, an above-market rate to toxic flow so that arbitrageurs are not able to capture as much of their profit.

Naturally, in a live setting, this is a dynamic rather than static game, and the behavior of traders will change as such measures are implemented. Nevertheless, it is still useful to perform an initial characterization of trade flows and wallets, which will give us a sense of what measures could be implemented to differentiate various sources of flow and their toxicity levels.

Price discrimination

As before, we analyze the profitability of the Uniswap ETH/USDC liquidity pools, using 5-minute markout windows calculated using Binance ETH/BUSD spot data. We are (naturally) able to reproduce our initial result showing that LP PnL drops off sharply as notional swap size in USD increases:

Interestingly, notice that most of the pools' losses come from swaps with a large notional size. Even though most notional swap sizes correspond to positive average PnL (for the pool), because it is the larger notional swap sizes that correspond to negative average PnLs, liquidity providers' PnL ends up negative overall. (Recall here that positive PnL means that the swapper is losing money and the liquidity pool is profiting, and vice versa.)

This initial observation motivates the idea of price discrimination: what if the liquidity pool could preferentially charge higher swap fees to incoming swaps that are more likely to be toxic flow (i.e., result in future negative PnL markouts for the liquidity pool)? One simple example would be to look at the notional size of each swap and increase the fee enough so that, according to the statistics above, the expected PnL of the swap is at least zero. This would result in a substantially higher and net positive PnL for liquidity providers!

However, this has an obvious problem: incoming swaps can be split up over multiple swaps or contract calls, so if it becomes known that the notional size of the swap is the determining criterion for the application of a supplemental swap fee, then swappers will be able to easily "dodge" this fee. Of course, one might imagine that more sophisticated mechanisms could be applied to try to catch these cases, which will then motivate yet more sophisticated evasion mechanisms… In general, once the mechanism of price discrimination is introduced, we begin to participate in an adversarial, "cat-and-mouse" game where sophisticated arbitrageurs seek to mask their presence from would-be discriminators.

This is not so different from the situation in traditional markets, where market makers seek at all costs to characterize toxic flow and charge them an appropriately higher spread, whereas proprietary firms with meaningful alpha will try to mask their presence in the markets to improve order execution! Indeed, in some sense, it is the AMM setting which is unusual, in that it currently has no mechanisms to allow for variation in the fee rate depending on the originator of the swap.

We will first focus on characterizing toxic vs. nontoxic flow, reserving a more detailed discussion of how price discrimination might be best implemented for a future post. Beyond notional swap size, a natural question to ask is whether or not the "freshness" of a wallet — the number of swaps previously observed for that wallet in the ETH/USDC liquidity pool — is a predictor of the toxicity of an incoming swap.

Again grouping by notional swap size, we find that swaps with small notional size tend to originate from wallets with fewer overall swaps, whereas swaps with larger notional size — also more likely to result in negative pool PnL — originate from

wallets with a much more extensive swap history:

This is unsurprising from an intuitive standpoint: swappers who only swap several dollars' worth of ETH probably do not do so very often, whereas arbitrageurs swapping 6 figures back and forth have no reason (for now) to switch wallets with any frequency.

We can look more directly at the relationship between the number of previously observed swaps for the originator of an incoming swap versus the actual markout PnL of that swap:

Doing so, we see that wallets with relatively limited swap history are indeed likely to give rise to swaps that are profitable for the liquidity pool, i.e., their swaps are likely to constitute nontoxic flow. It is only when we see a swap history of ≥ 40 swaps that the average markout PnL begins to decline, reaching negative PnL for swaps coming from wallets with ≥ 500 previous ETH/USDC swaps.

Curiously, we see that for wallets with quite extensive ETH/USDC swap histories, the expected PnL returns to positive levels. Why might this be the case? If we examine the data to see which wallets have originated ≥ 10,000 swaps in the ETH/USDC pool, we find that there are only fourteen such wallets! These fourteen wallets are composed of:

- One wallet, 0x1fd34033240c95aabf73e186a94b9576c6dab81b, which seems to focus on cyclic arbitrage between Uniswap pools
- One wallet, 0xe92f359e6f05564849afa933ce8f62b8007a1d5d, which is a CoWSwap order solver
- Twelve other wallets all associated with the infamous MEV bot 0xa57

The first wallet, and the only non-0xa57-associated MEV bot, has an average PnL of +5.58 basis points in the ETH/USDC pool, perhaps because in order to take advantage of price dislocations on other Uniswap pools, they effectively end up as a relatively insensitive taker of whatever ETH/USDC prices happen to be at the time of the dislocation. Of course, this means that this arbitrageur is perhaps leaving a little bit of money on the table from poor ETH/USDC swap execution, suggesting that whatever edge they have will eventually decay in favor of actors willing to execute different arbitrage legs on different venues!

The second wallet, a CoWSwap order solver, presumably passes through a wide variety of orders from retail and non-retail traders alike, and has an average PnL of +3.93 basis points. Interestingly, the other twelve wallets have average PnLs in the ETH/USDC pool ranging from -3.47 to +2.21 basis points, all lower than the average PnLs of the non-0xa57-associated wallets and suggestive of a great deal of sophistication on the part of 0xa57.

One very clear takeaway is that when we have high-quality sources of nontoxic flow, such as the CoWSwap order solver, liquidity providers should actually be willing to give a discount off the usual fee rate to swaps originating from such addresses, much like how Robinhood gives retail traders to better spreads through Citadel! If the existence of such a discount becomes publicized, then one might imagine that arbitrageurs might then try to "hide" their swaps alongside other CoWSwap traders! However, this is not necessarily straightforward; they will then be at the mercy of the order solver, unable to bundle transactions together, and so on. Additionally, the PnL of "whitelisted" addresses could be continually monitored to check that the swap flow originating from such addresses remains sufficiently nontoxic.

Another takeaway is that "not all MEV bots are built equal!" In the case of 0x1fd, and perhaps other purely-atomic cyclic arbitrageurs as well, a history of performing such arbitrage may very well be a credible sign of nontoxic flow, as far as the ETH/USDC pool itself is concerned. (That being said, one might then want to charge such arbitrageurs a substantially higher fee rate elsewhere...)

Fresh wallets

A different way of thinking about the problem of price discrimination is, instead of determining which wallets require upcharging due to their toxic flow, figuring out which wallets can be given discounts due to their non-toxic flow (much like the CoWSwap order solver in the previous section). Indeed, one could imagine a relatively high swap fee applied to all swaps, with a steep discount given to those addresses which are most likely to be non-toxic.

A very natural candidate for such fee discounts would be swaps originating from fresh wallets with zero, or close to zero, swaps in the ETH/USDC pool. As we saw earlier, when a swap comes in from an address with very limited swap history, it is on average likely to be a positive-PnL trade for the liquidity pool. It is of course trivially possible for arbitrageurs and informed traders to shift assets to new wallets in order to "masquerade" as retail traders. At the same time, it is valuable to note that:

- Moving assets between wallets costs both time and money and incurs significant additional logistical complexity.
- Stricter criteria can be established for wallets to "qualify" as non-toxic flow; for example, a fee discount could be given to wallets that have swapped more than 10 times, but fewer than 1,000 times, in the ETH/USDC pool with average positive PnL for the liquidity pool. (Various other combinations of criteria can be envisaged.) In other words, it can be made prohibitively hard for toxic flow to "pretend" to be non-toxic flow with sufficiently sophisticated checks.

Previously, we observed that the expected PnL of a swap varies based on both the notional swap size and the amount of swap history for its originating address. Can we say anything about how they co-vary? Below, we separated swaps into different colored groups depending on how many swaps had previously been observed for a given swap's originating

address, then plotted the relationship between notional swap size and PnL for that subgroup specifically:

Consistent with our intuitions, it appears that the bulk of toxic flow (1) is associated with high notional swap size and (2) originates from a consistent set of addresses that repeatedly trade in the ETH/USDC pool. When the notional size of a swap is low, or when it originates from a relatively fresh address, it is generally a profitable trade for the liquidity pool.

It is actually quite remarkable that for relatively “fresh” addresses, as the notional swap size increases, the expected PnL of the swap (for the liquidity pool) also increases! In fact, for completely fresh wallets with zero ETH/USDC swap history, swaps that are over the 80th percentile in notional swap size are usually profitable for the liquidity pool to the tune of over 10 basis points — quite a substantial margin of profitability.

Recall that in our original analysis of notional swap size versus swap PnL, we observed that at the far right end of notional swap size, the PnL began to noticeably increase. Here, we have cleanly shown that the downward PnL trend continues when we restrict to swaps originating from frequent traders, and that the upward PnL trend at the right end is completely associated with relatively fresher wallets. This is plausibly due to a combination of illicit hacking proceeds being swapped to ETH in a highly time-sensitive, price-insensitive manner as well as a sort of “retail whale” phenomenon where people who like to swing around large amounts of money on occasion are especially poor traders.

As a brief aside, we might wonder how we can tell whether or not those large swaps from new wallets are, in fact, largely from protocol exploiters! It is difficult to say without directly examining the swapping addresses themselves. However, we might argue that protocol hackers have a natural preference to swap into ETH, which can be anonymized and mixed. Therefore, if we look at the ETH buy-versus-sell imbalance, we should expect to see the imbalance tilt toward the “buy” direction as notional swap size increases (among swaps originating from relatively fresh addresses).

If we actually plot this metric for swaps originating from addresses with 10 or fewer prior ETH/USDC swaps, we find that there is a slight tilt in the buy-sell imbalance when notional swap size exceeds the 90th percentile:

In contrast, if we plot the buy-sell imbalance for swaps originating from addresses with 50 or more prior ETH/USDC swaps, we find the opposite effect, namely, a trend toward selling ETH when notional swap size exceeds the 90th percentile:

Although hardly a “smoking gun,” these observations are consistent with insensitive “exploiter flow” contributing to the rise in PnL as notional swap size exceeds the 90th percentile. That being said, it should be noted that the absolute levels of buy-sell imbalance are still relatively muted, so even if this “exploiter flow” has a detectable effect, it is only one of multiple contributing factors to the whole picture.

Separately, we might wonder about the persistence of a a given wallet’s PnL over time. If a fresh wallet’s very first swap has a high or low markout PnL, how well does that predict their markout PnL over all future observed swaps?

Generally speaking, while we do see a positive correlation between initial and average PnL, the relationship is quite weak. If we exclude all users with fewer than 20 swaps (so that the initial swap PnL does not excessively bias the calculation of the average PnL), there is only a weak positive correlation between initial and average PnL to the tune of $r=0.03$.

Interestingly, it is difficult to find a subset of swappers for whom there is a clear and consistent correlation between the initial swap PnL and their average swap PnL. If there are indeed arbitrageurs who come into the pool and immediately begin repeatedly swapping at large size and at prices unfavorable to the pool, we ought to be able to identify them as a subset of swappers for whom initial and average swap profitability are quite well correlated. However, even if we restrict to the subset of swappers who come in with a notional size above $1,000,000, the correlation between initial and average PnL only increases to a modest $r=0.11$, and most of the relationship appears to be coming from swappers with positive PnL who are themselves losing money relative to Binance markout prices!

One can imagine various reasons why this prevails; for example, perhaps profitable traders and arbitrageurs are trading multiple strategies from the same wallet, only a subset of which might be expected to be unprofitable for the liquidity pool. In this situation, although such a wallet might frequently originate large swaps with deeply negative PnL (for the pool), it may not necessarily do so consistently.

Autocorrelation of wallet performance

Stepping back a little bit, we may recall that we struggled to find a strong correlation between a wallet’s first swap and its average PnL over all swaps. One natural explanation may be that even for informed traders, individual swaps are quite noisy, and we may find stronger “persistence” of a wallet’s performance over time if we aggregate performance over larger windows. (For example, when analyzing the performance of mutual funds, the finance literature typically aggregates over monthly or multi-month intervals.)

Following this line of thought, we aggregated wallets’ swaps into chronological groups of 50 swaps each (restricting to wallets with at least 250 swaps) and calculated the average PnL within each group. (We chose a fixed number of swaps per group rather than chronological grouping because we felt it was plausible that swap activity, even among systematic traders, might vary substantially from one month to another.) We expect that an informed trader who consistently makes profitable trades against Uniswap liquidity will exhibit high autocorrelation of the sign of their PnL across consecutive “buckets” of swaps.

At an initial glance, there appears to be no obvious relationship between the average swap PnL and the autocorrelation of the signs of a given wallet's PnL:

However, if we look closer, we notice that there is a curious subset of wallets who always have negative PnL in every single group of swaps.

Previously, we found it illuminating to segment swaps by their notional size, observing that large swaps are more likely to be toxic flow (negative pool PnL). We take the same approach here by plotting a wallet's average notional swap size against its average PnL and coloring each point (corresponding to a single wallet) according to its PnL autocorrelation across groups of swaps:

There is a surprising amount of structure in this plot! Notably, there are three major clusters of wallets:

1. One cluster of positive PnL wallets (unprofitable traders) with relatively low notional swap sizes
2. Two clusters of negative PnL wallets (informed or toxic flow) with high notional swap sizes, one of which has extremely high PnL autocorrelation

The two negative-PnL clusters likely constitute a fairly high proportion of the toxic flow to which ETH/USDC LPs are subject, with the high-autocorrelation cluster constituting a qualitatively different "type" of arbitrageur from the lower-autocorrelation cluster.

What proportion of ETH/USDC LPs' overall losses come from these "clusters" of wallets associated with toxic flow?

- The smaller of the negative-PnL clusters with very high autocorrelation is made up of 52 wallets responsible for aggregate pool PnL of -23 million USD
- The larger of the negative-PnL clusters with lower autocorrelation is made up of 312 wallets responsible for aggregate pool PnL of -124 million USD

Striking numbers in light of the fact that the total ETH/USDC pool PnL is merely -42 million USD, meaning that outside of this subset of 364 wallets — i.e, the other 454,091 wallets in the dataset — are in fact immensely profitable for Uniswap LPs to the tune of positive 104 million USD! These results underscore, more than ever before, the importance of order flow segmentation and price discrimination to sustainable AMM liquidity provisioning.

Conclusion

To briefly summarize our findings so far:

- Fresh wallets are more likely to originate small, positive-PnL swaps, with toxic flow largely coming from very frequent and large swappers
- In fact, for wallets with limited swap history, expected pool PnL actually increases with notional swap size, and this is probably a combination of insensitive "exploit flow" for ETH purchases and "retail whales"
- The PnL of a wallet's first swap is, in general, not a strong predictor of its average PnL across all swaps; however, aggregating across larger groups of swaps, we can clearly identify "clusters" of wallets associated with toxic flow
- There is a small subset of wallets responsible for originating a large quantity of toxic flow, generating losses on the order of 150 million USD for Uniswap LPs

Although segregation of toxic flow will prove to be challenging, especially as traders adapt their methods to become less detectable, we have at least made some initial steps here toward characterizing the different sources of swap flow and their properties.

In the next post in this series, we will explore properties of toxic flow, such as the consistency of their PnL or the timeframe of their alphas, in more detail. We will also sketch out workable mechanisms for price discrimination that overcome the challenges of working within the EVM.

-0xfbifemboy