Protocol Fees and Liquidity Incentives in the 0x Protocol

by Amir Bandeali, Will Warren, Weijie Wu, and Peter Zeitz*

April 11, 2019

1 Abstract

We are proposing the introduction of an ecosystem fee on the taker side of each 0x trade. The protocol fee will be payable in ETH and the fee amount will be pegged to the gas fee takers pay for inclusion of their transactions in blocks. A portion of the fee revenues generated will be allocated to the 0x community treasury to fund ongoing development of the 0x ecosystem. The remaining fee revenues will be paid out to 0x market makers as a liquidity rebate. The rebate scheme rewards market makers in proportion to a) the amount of liquidity they provide and b) their stake of ZRX tokens. This paper provides a detailed analysis of the fee and rebate scheme's economic effects on makers, takers, and ZRX holders.

*Corresponding Author:peter@0x.org

Contents

2 Introduction

2.1 The 0x Protocol Fee

We are proposing the introduction of an ecosystem fee on the taker side of each 0x trade. The protocol fee will be payable in ETH and the fee amount will be pegged to the gas fee takers pay for inclusion of their transactions in blocks. For example, a taker who currently pays a USD $0.10 gas fee in ETH would now be required to attach an additional USD $0.10 in ETH as a 0x protocol fee. Gas prices are extremely variable across transactions and this variability allows for pricing flexibility among users who do not require immediate settlement. A patient taker could submit trades at a gas cost of only USD $0.01 and would then pay a protocol fee of only USD $0.01. In some situations, it is not possible for takers to fill orders without offering miners substantial payments to prioritize their requests over those of competing takers. Arbitrage bots, for example, compete for rights to fill profitable orders through bidding wars where they offer miners larger and larger gas prices to prioritize their transactions. Under this proposal, arbitrage bots would need to bid up the 0x protocol fee in order to obtain these fills. In effect, the system would recover some the revenue currently paid to ETH miners as a 0x protocol fee. These revenues could then fund 0x community projects or simply be returned to the market maker as a liquidity rebate.

2.2 Why Introduce a Protocol Fee?

We have three major motivations in proposing this new fee scheme. Firstly, in the long-term, sustaining development of the 0x protocol will require a source of ongoing revenue. In the future, a portion of fee revenue will be set aside to fund a 0x community treasury. The 0x community treasuty will be an on-chain reserve of ERC20 tokens that funds public goods that benefit the 0x community, inclusive of both ongoing work on protocol development and grants for third-party projects that make use of 0x exchange functionality. Decisions on how best to make use of funds stored in the 0x community treasury will be managed through a token voting scheme. A second motivation of the fee scheme is to create incentives for ZRX ownership among liquidity providers. We feel that alignment of interests between liquidity providers and the 0x protocol is critical for the project's long-term success. Smart contracts will rebate a portion of the fee revenue collected to market makers who maintain a ZRX stake as a liquidity incentive. The ability to earn these rebates will encourage greater participation of market makers in ZRX ownership. A third motivation for the fee scheme is to create a dynamic element to limit order pricing, so as to reduce inefficiencies associated with on-chain trade settlement. At present, the high latency of order processing on decentralized exchanges exposes market makers to an increased risk of incurring losses from arbitrage activity. The fee scheme will recover a portion of arbitrage bots' gas expenditures as a 0x protocol fee. Since the recovered funds are rebated to market makers, 0x liquidity providers will ultimately benefit from price improvement.

2.3 Design Goals for the Protocol Fee and Rebate Program

We had several goals for our fee and rebate program and these guided us towards a specific choice of a fee and rebate schedule. Firstly, we wanted to levy higher fees on trades that impose negative externalities on third party users of the 0x protocol.[1] Liquidity suppliers lose money in arbitrage trades and must pass on these losses to liquidity demanders through increases in the bid-ask spread. Our fee scheme dynamically increases a limit order's asking price in the event that multiple takers compete for a fill. Since arbitrage trades are charged much larger fees relative to other types of trades, this serves as a type of Pigovian tax on arbitrage activity. A rebate of these protocol fees to market makers will reduce their arbitrage losses and allow them to charge liquidity demanders lower prices.

A second goal of our fee and rebate program is to maintain a level playing field for small and large market making firms. At most crypto-exchanges, liquidity rebate schemes offer significant discounts to high volume traders and/or large-scale trading firms who can amortize a large ownership stake over a large number of trades. We see these programs as unfair because they put preferential fees out of reach for small trading firms. Our proposed fee scheme allows both small- and large-scale market makers to access the same fee and rebate structure. This should help to preserve healthy and fair competition among market makers using the 0x protocol. In the long-term, fairness will be important to avoiding conflicts between groups of 0x users with competing interests.

A third goal of our fee scheme is to create strong incentives for user ownership. As we introduce token voting mechanisms such as the recently launched ZEIP process, ZRX owners will gain increasing control over the 0x protocol's long-term development. To ensure that the token voting process empowers the protocol's user community, we must create incentives for users to maintain a ZRX ownership stake. To incentivize user ownership, liquidity providers will be required to maintain a ZRX stake proportional to their usage in order to obtain liquidity rebates. If only a subset of market makers choose to participate in staking, this group will be able to capture the entire pool of 0x protocol fees.

A fourth goal of our fee scheme is to define a process for determining rebate payouts that is simple, intuitive, and resistant to gaming. Ideally, market makers participating in the scheme should be able to refer to simple rules of thumb to determine how much ZRX to stake in order to maximize their business income. These rules should be forgiving, so that staking a little more or a little less than the optimal amount would have a negligible impact on market maker profitability. Finally, the fee and rebate scheme should ensure that the distribution of rebate income remains consistent over time even as market conditions change. Keeping rebate payouts simple, inutitive, and predictable will make it much easier for market makers to evaluate how staking would affect their business.

3 From Design Goals to a Specific Choice of Rebate Function

The mathematical formula determining the distribution of rebate payments is somewhat involved, so it is useful to start out with a high-level overview and a few mathematical definitions. Firstly, in order to reduce gas costs associated with administration of the rebate program, all user fees will be pooled, escrowed in a smart contract, and then paid out as lump sum liquidity rebates once every three months. In the event that surplus fees are left over after paying rebates, these fees will be sent to the 0x community treasury. This can

[1]For an in depth discussion of negative externalities associated with arbitrage trades, see the 0x technical paper "Adding a Selective Delay to Decentralized Exchange."

happen, for example, if some market makers do not maintain a stake sufficient to claim rebates associated with their own trading activity. Let $\tau_t \geq 0$ denote the total amount of user fees collected during epoch t. Furthermore, let ρ_t denote an additional amount removed from the fee pool to fund the 0x community treasury (in which case $\rho_t < 0$) or alternatively added to the fee pool to incentivize additional liquidity provision (in which case $\rho_t > 0$). The amount of funds available for distribution as rebates at the end of epoch t is given by $(\tau_t + \rho_t)$.

The next element of the formula is a function that determines the share of the rebate pool payable to each market maker. This formula considers two types of resources that liquidity providers contribute to the 0x protocol and incentivizes each market maker to contribute both resources in equal proportion. The first resource is ZRX stake. At the beginning of each three month epoch, ZRX holders who wish to participate in the rebate program must commit a quantity of ZRX stake to a time-locked deposit. The stake contributor can then either allocate this ZRX stake to his own trading account or rent out the use of this ZRX stake to a third party. Let $z_{it} \in [0,1]$ denote the ratio of market maker i's stake relative to the total stake committed by all market makers during epoch t. We refer to z_{it} as maker i's staking share. The second resource is the market maker's personal contribution to protocol fees. Whenever market makers complete trades, a smart contract will increment a running tally of protocol fees attributed to their staking addresses. Let $\tau_{it} \in [0,1]$ denote the proportion of all fees collected in epoch t that are assigned to maker i's staking address. We refer to τ_{it} as maker i's fee share.[2] Finally, a reward function, $s(z_{it}, \tau_{it})$, aggregates these two measures in order to compute a percent share of the epoch t rebate pool payable to market maker i.

3.1 Homotheticity Property

In order to meet our design objectives, we require the function $s(z_{it}, \tau_{it})$ to exhibit a specific set of mathematical properties. Firstly, we would like the ability to earn rebate awards to be invariant to the scale of a market-making firm. In particular, independent market makers should not be able to increase their awards by merging trading activity under a single account or by subdividing their activity across multiple accounts. To ensure scale invariance, $s(z_{it}, \tau_{it})$ should satisfy $cs(z_{it}, \tau_{it}) = s(cz_{it}, c\tau_{it})$ where c is any positive real number, $c \in \mathbb{R}^+$. A function that has this property is said to be homothetic of degree one. Suppose that we set c equal to $\frac{1}{\tau_{it}}$. Homotheticity implies that we can rewrite the function $s(z_{it}, \tau_{it})$ as shown in Equation 1.

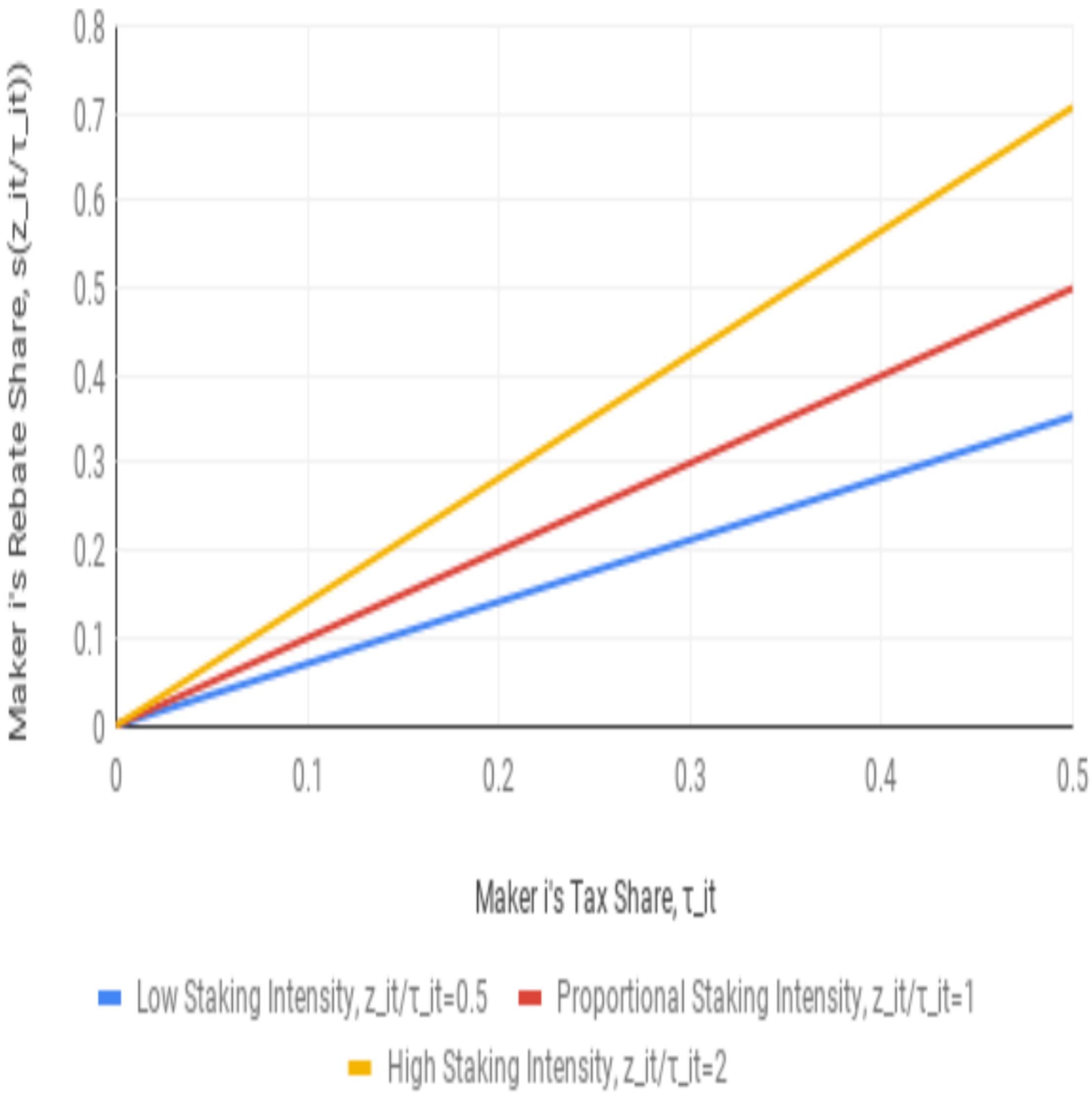
$$s(z_{it}, \tau_{it}) = \tau_{it} s\left(\frac{z_{it}}{\tau_{it}}\right). \quad (1)$$

In Equation 1, the ratio $\frac{z_{it}}{\tau_{it}}$ indicates the ratio of market maker i's ZRX stake to his fee generating activity. We refer to $\frac{z_{it}}{\tau_{it}}$ as market maker i's staking intensity. Likewise, the multiplicative term τ_{it} captures the market share of market maker i. We can think of τ_{it} as a measure of the scale of market maker i's trading activity. As shown in Figure 1,

[2]Note that τ_{it} is the fraction of τ_t contributed by maker i. The amount of fees contributed by maker i is thus given by the product, $\tau_{it}\tau_t$.

homotheticity implies that for any fixed staking intensity, $\frac{z_{it}}{\tau_{it}}$, rebate payments scale linearly with the number of trades a market maker performs.

Figure 1: Rebate Payments Scale Linearly with Trading Activity



The x-axis is maker i's share in total fee generation, τ_i, and the y-axis is the proportion of the liquidity rebate pool paid to maker i, $s(z_{it}, \tau_{it})$. Makers who maintain a higher staking intensity (yellow line) recover a larger proportion of the fees they generate as liquidity rebates. For any given choice of staking intensity, rebate payments scale linearly with the volume of their trading activity.

3.2 Concavity Property

A second aim of the program is to encourage market makers to contribute stake and fee resources in equal proportion. In other words, we would like all market makers to target an ownership stake proportional to their usage. This obtains if all market makers to target a staking intensity of $\frac{z_{it}}{\tau_{it}} = 1$. To get an intution for how this objective can be mapped into a specific mathematical property, suppose the following:

1) There are two market makers i and j, whose fee shares are given by $\tau_{it} > 0$ and $\tau_{jt} > 0$, respectively. Let λ equal the ratio of the fee revenue contributed by maker i to the total fee revenue contributed by both makers, $\lambda = \frac{\tau_{it}}{\tau_{it}+\tau_{jt}}$.

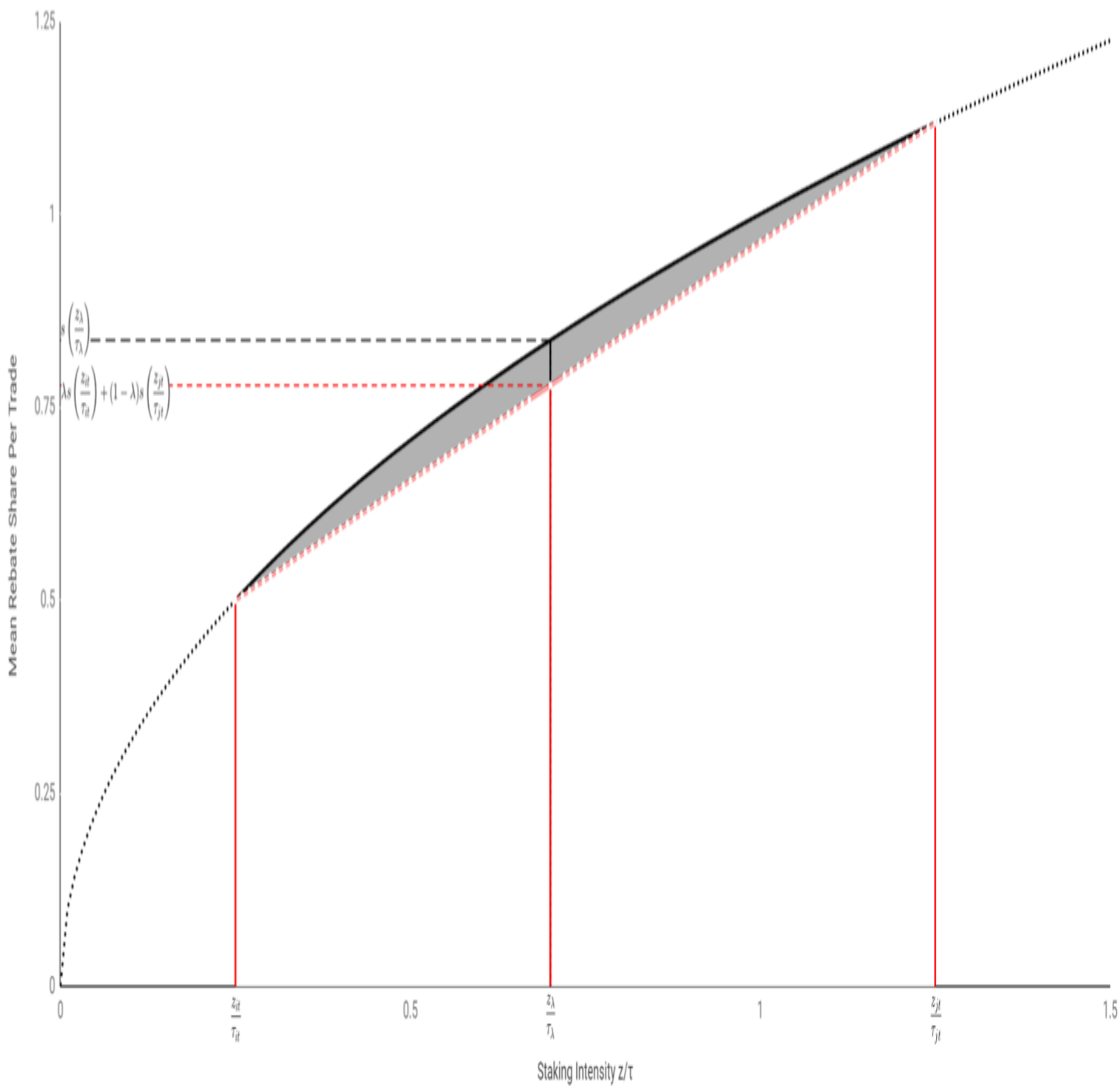
2) The two market makers stake with different intensities $\frac{z_{it}}{\tau_{it}} \neq \frac{z_{jt}}{\tau_{jt}}$.

We would like our rebate function to encourage the two market makers to trade ZRX with one another, until they reach a point where both makers stake with equal intensity. The mathematical property that assures this is known as strict concavity. A function $s\left(\frac{z_{it}}{\tau_{it}}\right)$ is said to be strictly concave iff for any two staking intensities $\frac{z_{it}}{\tau_{it}}$ and $\frac{z_{jt}}{\tau_{jt}}$ and for any weighted average of these intensities, $\frac{z_\lambda}{\tau_\lambda} = \lambda \frac{z_{it}}{\tau_{it}} + (1-\lambda)\frac{z_{jt}}{\tau_{jt}}$, where $0 < \lambda < 1$, the inequality shown in Equation 2 holds.

$$s\left(\frac{z_\lambda}{\tau_\lambda}\right) \geq \lambda s\left(\frac{z_{it}}{\tau_{it}}\right) + (1-\lambda)\left(\frac{z_{jt}}{\tau_{jt}}\right) \text{ with equality iff } \frac{z_{it}}{\tau_{it}} = \frac{z_{jt}}{\tau_{jt}}. \quad (2)$$

As plotted in Figure 2, concavity implies that whenever $\frac{z_{it}}{\tau_{it}} \neq \frac{z_{jt}}{\tau_{jt}}$, market makers i and j will be able to increase their rebate income by trading stake with one another until they reach a point where they both stake with equal intensity, i.e. $\frac{z_{it}}{\tau_{it}} = \frac{z_{jt}}{\tau_{jt}} = \frac{z_\lambda}{\tau_\lambda}$.

Figure 2: Liquidity Providers Benefit From Trading Stake until both Stake with Equal Intensity



The x-axis indicates staking intensity and the y-axis indicates the proportion of fees on each trade that are returned to the maker as a liquidity rebate. In the figure, maker i selects a lower staking intensity $\frac{z_{it}}{\tau_{it}}$ and is rebated a smaller proportion of the fees he generates, whereas maker j stakes with a higher intensity $\frac{z_{jt}}{\tau_{jt}}$ and is rebated a larger proportion. Averaging across the two makers, the mean proportion of fees they recover as a rebate is given by the dashed red line, $\lambda \mathrm{s}\left(\frac{z_{it}}{\tau_{it}}\right)+(1-\lambda)\mathrm{s}\left(\frac{z_{jt}}{\tau_{jt}}\right)$. If the two makers exchanged stake until they reached a point where both stake with equal intensity, then they could recover a larger mean rebate payment $s\left(\frac{z_\lambda}{\tau_\lambda}\right)$ using the same resources.

Combining the two properties of homotheticity and concavity leads to an interesting result. Suppose that all market makers are identical except for the total amount of working capital they have available to stake and perform trades. Let $\left(\frac{z_{it}}{\tau_{it}}\right)^*$ denote the optimal staking intensity for some market maker i. Under any reward function that is homothetic of degree one [Equation 1] and concave [Equation 2], all market makers will stake with the same intensity in market equilibrium. In order for all market makers to select the same staking intensity, we must have that $\frac{z_{it}}{\tau_{it}}^* = 1$ for all i. Suppose further that that the subsidy is zero, $\rho_t = 0$, and that the reward function is scaled such that 100% of fees collected are paid out as rebates. This implies then that $s\left(\left(\frac{z_{it}}{\tau_{it}}\right)^*\right) = s(1) = 1$. The market maker's equilibrium payout can then be calculated as $\tau_t s\left(\tau_{it}\left(\frac{z_{it}}{\tau_{it}}\right)^*, \tau_{it}\right) = \tau_t \tau_{it} s\left(\left(\frac{z_{it}}{\tau_{it}}\right)^*\right) = \tau_t \tau_{it}$. In words, a maker who contributes $\tau_t \tau_{it}$ in protocol fees to the rebate pool and chooses the optimal staking intensity, $\left(\frac{z_{it}}{\tau_{it}}\right)^* = 1$ will recover exactly the amount he puts in, $\tau_t \tau_{it}$, as a liquidity rebate. We can see then that homotheticity and concavity result in a rebate scheme that satisfies a basic definition of fairness.

3.3 Stability of Input and Output Price Relationships

A third aim of the program is to ensure that decisions for market makers and ZRX holders are simple and intuitive and that payoff outcomes for these groups obtain are stable and predictable. These goals are more difficult to put in precise mathematical terms. Accordingly, we simply present the rebate function and describe some of its properties. Our hope is that readers will immediately perceive why these properties help to simplify decisions and ensure predictable and stable outcomes. To simplify exposition, we simply assert that the function has these properties here. Interested readers can consult the mathematical appendix for formal derivations.

This function we plan to use to divide rebates is known as a Cobb-Douglass production function, and it has the functional form shown in Equation 3. In Equation 3, market maker i's share of liquidity rebates is computed as a weighted geometric average of his fee share and stake share, where the parameter $\alpha \in (0, 1)$ is a weight placed on the fee share when computing this average. Likewise, $(1 - \alpha)$ is the weight placed on the stake share.

$$s(\tau_{it}, z_{it}) = \tau_{it}^{\alpha} z_{it}^{1-\alpha} \text{ where } 0 < \alpha < 1. \tag{3}$$

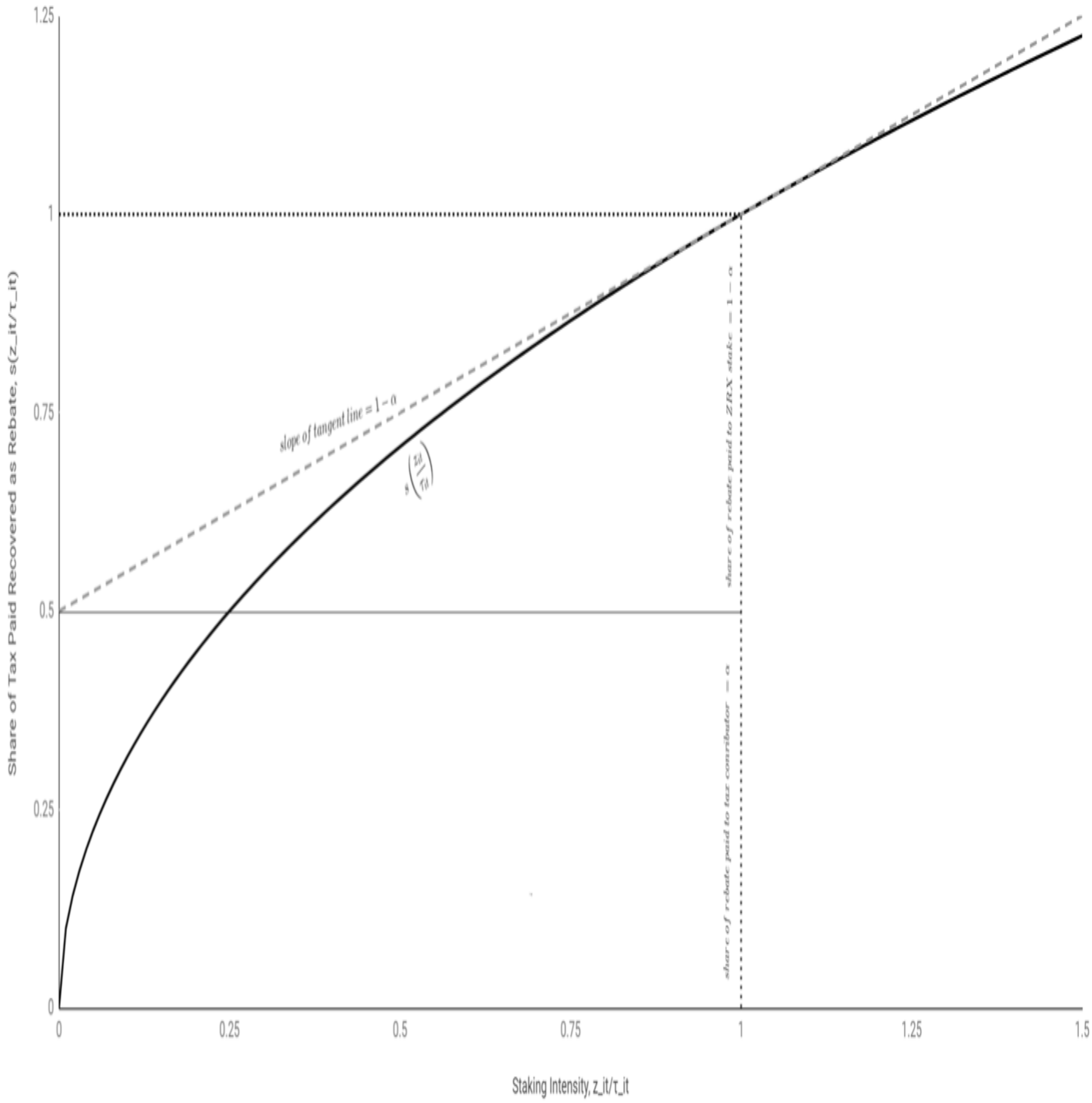
The Cobb-Douglas function is the most commonly used function in economic analysis and owes its popularity to its ability to ensure stable relationships between input and output prices. Frequently, two or more distinct inputs are purchased on a so-called input market. A firm then utilizes these inputs to create some type of output, the sale of which generates revenue. A major task of economic theory is to explain how firms' divide these revenues across multiple input suppliers. For example, if an iPhone sells for \$1000, why do markets allocate only \$10 per phone of this value to phone assembly workers, while Apple's shareholders receive \$500 per phone in profit?

In our situation, τ_{it} and z_{it} are the two distinct production inputs and the share of the rebate pool, $s(\tau_{it}, z_{it})$, is the output they generate. Suppose, for example that Bob creates maker offers that these maker offers generate τ_{it} in fees. Simultanesouly, Alice supplies the necessary stake z_{it} that allows Bob to claim rebates. We would like to understand how a competitive marketplace will divide rebate income between Alice and Bob. Intuitively, we can see that if Alice is granted too large a share of rebate income, then Bob will have incentive to purchase his own stake and trade on his own rather than rent stake from Alice. Likewise, if Bob is granted too large a share, than Alice will have an incentive to perform the trades herself rather than rent out stake to Bob. In market equilibrium, revenues are divided in such a way that Alice and Bob are indifferent between working together or striking out on their own.

The Cobb-Douglas function has the unique property that the division of revenue between Alice and Bob always remains constant regardless of any changes in external market conditions. In a competitive equilibrium, the individual who contributes the fee input will always receive a share α of the total rebates generated, where α is the exponent on the fee share in Equation 3. Conversely, the individual who contributes stake will always receive a share of $(1 - \alpha)$. Under all other known concave and homothetic reward functions, relationships between input and output prices depend on time-varying factors. For example, an uptick in the ZRX price could cause Alice to receive a smaller or larger share of the rebate income. Given the irrationality and unpredictability of the crypto market, this would introduce unwanted uncertainty into making and staking decisions.

Figure 3 illustrates the relationship between the parameter α and the percentage of rebates paid to ZRX holders. A higher value of α reduces the weight on stake contributions in the determination of rebate payouts and increases the weight on fee contributions. As can be seen from the figure, the proportion of rebate income paid to ZRX holders is given by $1 - \alpha$. Likewise, the equilibrium price charged for renting one unit of ZRX stake for one epoch is given by $(1 - \alpha) \frac{\tau_t + \rho_t}{z_t}$,where τ_t is the total amount of fee revenue collected, ρ_t is a net fee or net subsidy, and z_t is the total number of units of ZRX staked.

Figure 3: Equilibrium Division of Rebate Income



The x-axis is staking intensity and the y-axis indicates the proportion of fees on each trade that are returned to the maker as a liquidity rebate. The marginal increase in rebate income a maker would obtain from increasing his staking intensity is given by the slope of the curve $s(\frac{z_{it}}{\tau_{it}})$. We know that all market makers must select a staking intensity of $\left(\frac{z_{it}}{\tau_{it}}\right)^* = 1$ in equilibrium. For the Cobb-Douglas function, the slope of $s(\frac{z_{it}}{\tau_{it}})$ evaluated at $\left(\frac{z_{it}}{\tau_{it}}\right)^* = 1$ is equal to $1 - \alpha$. This slope determines the share of the rebate pool paid to stakeholders and in turn the price of renting one unit of ZRX. Suppose that all of the fees generated are returned as rebates, so that $\rho_t = 0$ and the total value of the rebate pool is τ_t. Suppose further that z_t units of ZRX are staked. The equilibrium price of renting one unit of ZRX stake is then $(1 - \alpha)\,\frac{\tau_t}{z_t}$.

4 Effects of the Fee and Rebate Model on Key 0x Project Stakeholders

4.1 Discounted Present Value of ZRX Holder Income

One benefit of the Cobb-Douglas function is that it makes quantification of the benefits of ZRX ownership simple. Under this function, ZRX holders receive a fraction $(1-\alpha)$ of liquidity rebates in equilibrium. Given some projections about the value of current and future rebate pools and a risk-adjusted interest rate it is trivial to express this benefit in terms of its discounted present value. Suppose that based on some assumptions we compute that the discounted present value of all current and future liquidity rebates is X. The present value of the rebate stream accruing to ZRX token holders is then simply $(1-\alpha)\,X$.

4.2 Market Makers Staking Decisions

This simplicity carries over into a large number of related decision problems. For example, suppose that a market maker must divide the working capital he uses to trade on the 0x protocol between a ZRX stake and a token inventory. Furthermore, suppose that a market maker expects to earn a fraction c of his total income from liquidity rebates and a fraction $1-c$ of this income from charging a spread. Finally, let's assume further that the number of trades the market maker completes scales linearally with the amount of working capital he allocates to borrow a trading inventory. For, example, if our market maker locks up 10% of his working capital to rent a ZRX stake, then he will have only 90% of the initial amount left over for trading and will thus complete 10% fewer trades. Under these assumptions, a market maker seeking to maximize profit will devote a fraction $(1-\alpha)\,c$ of his working capital to renting ZRX stake and a fraction $1-(1-\alpha)\,c$ to borrowing a trading inventory. This simple decision rule holds irrespective of any changes in the market price of ZRX, the interest costs associated with renting ZRX stake or borrowing trading inventory, and any changes in the scale of the market maker's trading operation.

4.3 Effects on the Bid-Ask Spread Charged to Different Groups

The staking requirement will mean that some of the working capital market makers use to supply liquidity on 0x order books will become unavailable. This reduction in the supply of liquidity will need to be absorbed through some combination of a decrease in aggregate trading volume on 0x relayers and an increase in the average bid-ask spread paid by takers. While these effects may sound alarming, it is important to keep in mind that an increase in the mean bid-ask spread systemwide does not necessarily imply an increase in the mean spread charged to retail traders.

Under this program, the total bid-ask spread would now have two components. Firstly, there will be the quoted bid-ask spread which is fixed by the limit order itself. In the mathematical appendix, we show that the fixed component of the bid-ask spread is expected to decrease under the fee and rebate program. In addition to this, there will now be a variable component of the bid-ask spread associated with the protocol fee. The variable component of the spread will differ accross individuals and be higher for takers that require immediate trade settlement. For retail traders, it will be possible to reduce the vairable component by submitting a trade at a lower gas price. In particular, suppose that a user submits a transaction that pays a fraction $f < 1$ of the average gas price paid by other users. Based on the mathemetical model we present in the appendix, patient users will actually be charged a lower bid-ask spread overall provided that $f < \alpha$, where α is the rebate function parameter defined in Equation 3. Conversely, users seeking to aggressively arbitrage 0x orders will need to offer larger than average gas prices and will thus pay significantly higher spreads.

A similar logic applies to changes in overall trading volume as well. Since arbtirage-related trades are expected to face a higher spread, we would expect arbitrage-related trading volumes to fall. Conversely, since retail traders will now face a lower spread, we would expect retail trading volume to increase. If the program works as intended, both market makers and ordinary users will see a net benefit from the introduction of the fee and rebate program. Conversely, individuals seeking to perform aggressive arbitrage against 0x orders and Ethereum miners are much less likely to benefit from the introduction of this program.

5 Mathematical Appendix: General Equilbirum Model of the Cobb-Douglas Liquidity Rebate Program

5.1 Liquidity Demand Function

Let θ denote the mean maker profit generated by each trade, inclusive of both profit accruing to the maker and profit collected as a protocol fee. We can think of θ as the mean bid-ask spread charged to takers.

Let γ denote the number of trades a market maker is able to perform using one unit of inventory per epoch. If γ is low, then makers will have to queue to complete trades, which then implies that there will be more depth in the order book for any given level of trading volume. Thus, all else equal, takers benefit from lower levels of both θ and γ.

The mean profit a market maker generates per unit of trading inventory per epoch is then $\theta\gamma$.

Let $q_d(\theta\gamma)$ denote the quantity of trades demanded by takers as a function of the product $\gamma\theta$.

We assume a constant elasticity of demand system as shown in Equation 4, where $\mu > 0$ is the elasticity of demand with respect to the product $\theta\gamma$ and q_0 is a normalizing constant.

$$q_d(\theta\gamma) = q_0 (\theta\gamma)^{-\mu} \tag{4}$$

Note here that this model will only be able to determine the product $\theta\gamma$ and cannot pin down specific values for θ and γ. In a future version, we plan to enrich the model by allowing the maker to choose a specific values of θ and γ, so as to understand how the fee and rebate program will affect this choice. For the time being, however, omitting this aspect substantially simplifies analysis.

5.2 Defining the Rebate Program

The rebate program will directly affect the willingness of market makers to provide liquidity. In particular, market makers will face an additional capital cost associated with maintaining a ZRX stake.

We use two parameters to describe the rebate program.

Let α denote the weight placed on fee contributions in division of the rebate pool. Note that the weight placed on stake is simply $1 - \alpha$.

Let c denote the proportion of trading profit that is captured in the protocol fee. Note that since all protocol fees are paid out as rebates, this is also the proportion of trading profit paid out as a liquidity rebate.

For simplicity, we will assume that the liquidity subsidy ρ_t is equal to 0.

5.3 Liquidity Supply

Let r denote the interest a market maker pays to borrow one unit of token inventory in terms of USD. We assume that r is exogenously determined. This is reasonable as long as trade volumes on 0x remain too small to absorb a significant fraction of token inventories. Let w denote the interest a market maker pays to rent one unit of ZRX stake in terms of USD. To simplify notation, we define the full set of market parameters as $\Theta = (r, w, \alpha, c, \theta, \gamma)$. Let $q_s(\Theta)$ denote the quantity of trades supplied as a function of Θ. In equilibrium, we must have that $q_s(\Theta) = q_d(\theta\gamma)$.

5.4 Staking Market Equilibrium

For simplicity, we denominate units of ZRX in terms of a percentage of the entire circulating supply and assume that in equilibrium all circulating ZRX are locked up in stake. Let $z_d(\Theta)$ denote aggregate demand for ZRX as a function of Θ. The price of renting ZRX, w, is determined by the equilibrium condition $z_d(\Theta) = 1$.

5.5 Market Maker's Profit Function

We assume that individual market makers are price takers and are only differentiated by the amount of capital at their disposal. Accordingly, market makers take aggregate market conditions, $(r, w, \alpha, c, \theta, \gamma)$, as given when making decisions.

Let l_{it} denote the number of units of token inventory borrowed by market maker i.

Let z_{it} denote the number of units of ZRX stake borrowed by market maker i.

Let z_t denote the total units of ZRX staked by all market makers.

Note that market maker i's share in total fee contribution is given by $\frac{l_{it}}{l_t}$ and that his share in total stake contributions is given by $\frac{z_{it}}{z_t}$.

Let $\pi(l_{it}, z_{it}; l_t, z_t, \Theta)$ denote the profit a market maker i earns in epoch t as a function of his decision variables l_{it} and z_{it}, l_t and the external market conditions Θ. Then $\pi(l_{it}, z_{it}; l_t, z_t, \Theta)$ can be expressed as shown in Equation 5.

$$\pi(l_{it}, z_{it}; l_t, \Theta) = (c\gamma\theta l_t)\left(\frac{l_{it}}{l_t}\right)^{\alpha}\left(\frac{z_{it}}{z_t}\right)^{1-\alpha} + (1-c)\,\gamma\theta l_{it} - r l_{it} - w z_{it} \tag{5}$$

5.6 Market Maker's Budget Constraint

Let k_{it} denote the amount of capital available to market maker i. A market maker allocation (l_{it}, z_{it}) is feasible if it satisfies the budget constraint shown in Equation 6 and the non-negativity constraints shown in Equations 7 and 8.

$$rl_{it} + wz_{it} \leq k_{it} \tag{6}$$

$$l_{it} \geq 0 \tag{7}$$

$$z_{it} \geq 0 \tag{8}$$

5.7 Market Maker's Decision Problem

The market maker's objective is to choose an allocation $\left(l_{it}^{*}, z_{it}^{*}\right)$ that maximizes profit given his budget constraint, as shown in Equation 9.

$$(l_{it}^{*}, z_{it}^{*}) = \arg\max \pi(l_{it}, z_{it}; l_t, z_t, \Theta) : rl_{it} + wz_{it} \leq k_{it} \tag{9}$$

Note that since $\pi(l_{it}, z_{it}; l_t, z_t, \Theta)$ is concave and homothetic, any interior solution where $l_{it}^{*} > 0$ and $z_{it}^{*} > 0$ must satisfy the first order condition shown in Equation 10.

$$\underbrace{\frac{1}{r}\left(\alpha\left(c\gamma\theta\right)\left(\frac{l_t}{z_t}\right)^{1-\alpha}\left(\frac{z_{it}}{l_{it}}\right)^{1-\alpha} + \left(1-c\right)\gamma\theta\right)}_{\text{Marginal Return to 1 USD investment in } l_{it}} = \underbrace{\frac{1}{w}\left(\left(1-\alpha\right)\left(c\gamma\theta\right)\left(\frac{l_t}{z_t}\right)^{1-\alpha}\left(\frac{z_{it}}{l_{it}}\right)^{-\alpha}\right)}_{\text{Marginal Return to 1 USD investment in } z_{it}} \tag{10}$$

5.8 Supply Side Equilbirum

In equilibrium, Equation 10 must be satisfied for all market makers. This then implies that $\frac{z_{it}}{l_{it}} = \frac{z_t}{l_t}$.

Substituting this condition into Equation 10 yields the expression for the ratio of the market makers' USD expenditure on renting stake to that on borrowing token inventories as shown in Equation 11.

$$\frac{wz_t}{rl_t} = \frac{(1-\alpha)\,c}{\alpha c + (1-c)} \tag{11}$$

Equation 11 says that the optimal division of the market maker's working capital depends solely on the two parameters we choose for the rebate program α and c. Note that this holds regardless of demand conditions, the prices r and w, or any other type of external market condition. The Cobb-Douglas function is the only known rebate distribution function that exhibits this unique property.

We can substitute Equation 11, $\frac{z_{it}}{l_{it}} = \frac{z_t}{l_t}$, and the market maker's budget constraint into Equation 5 to recover the rate of return the market maker earns on capital as shown in Equation 12.

$$\frac{\pi(k_{it};l_t,\Theta)}{k_{it}} = \frac{(\gamma\theta)\,(\alpha c + (1-c)) - r}{r} \tag{12}$$

We assume that market making is perfectly competitive and thus that new market makers will enter if the profit rate is greater than 0 and exit if the profit rate is less than 0. In equilibrium, the profit rate will be 0. This then implies the supply side equilibrium (Zero Profit) condition shown in Equation 13.

$$\begin{array}{cccc} (\alpha c + (1-c)) & \gamma\theta & = & r \\ \text{MM Share of USD Profit *} & \text{USD Profit Per Inventory Unit} & = & \text{USD Cost Inventory Unit} \end{array} \tag{13}$$

Note that under perfect competition, the supply curve is assumed to be horizontal (perfectly elastic). Under this market structure assumption, the supply curve is sufficient to determine the revenue generate by one unit of trading inventory in each epoch, $(\theta\gamma)$. To generate an upward sloping supply curve, we could simply assume that the interest rate r is increasing in the aggregate number of tokens borrowed, l_t. The right hand side of Equation 13 would then become an increasing function of r.

5.9 General Equilbirum

To compute the equilibrium number of trades we can simply substitute the condition for supply-side equilibrium, Equation 13, into the demand function in Equation 4. This yields the expression for the equilibrium number of trades shown in Equation 14.

$$q_d\,(\theta\gamma) = \left(\frac{r}{\alpha c + (1-c)}\right)^{-\eta} q_0 \tag{14}$$

Equation 14 is easier to interpret when expressed in terms of logs as shown in Equation 15.

$$\ln q_d\,(\theta\gamma) = \ln q_0 - \eta \ln r + \eta \ln(\alpha c + (1-c)) \tag{15}$$

In Equation 15, the term $(\alpha c + (1-c))$ is the proportion of trading profit paid out on the basis of trading activity, whereas $(1-\alpha)\,c$ is the proportion of trading profit paid out the basis of staking contributions. We can think of $(1-\alpha)\,c$ as the protocol fee levied on trading profits to support the 0x protocol. Given a demand elasticity η, Equation 15 says that this fee would be predicted to reduce the number of trades on 0x relayers by $(100 * (1-\alpha)\,c\eta)$%.

Part of any reduction in trading volumes would be absorbed through the reallocation of $100 * (1-\alpha)\,c$% of market makers working capital to stake. In addition to this, the fee could also provoke entry or exit among market making firms. If the demand elasticity is equal to $\eta = 1$, then reallocation of token inventories would be sufficient to absorb the reduction in trading activity and no market making capital would enter or leave the system. If demand elasticity is $\eta > 1$ (as is the case for a firm in a competitive industry), fees would cause $100 * (\eta - 1)\,(1-\alpha)\,c$% of market making capital to exit the system. In the interim epoch before this exit occurred, the fee would reduce the profit rates of incumbent market makers. If the demand elasticity is less than 1 (as is the case for a monopoly), fees would cause a $100 * (1-\eta)\,(1-\alpha)\,c$% increase in market making capital. In the interim epoch before this capital entered, the fee would increase the profit rates of incumbent market makers.

5.10 Heterogeneous Effects on Bid Ask Spread Charged to Different User Types

To completely model heterogeneous effects on various users types, we need to enrich the model so as to explain a market maker's decision to quote a particular mean bid-ask spread θ and obtain a given level of inventory turnover γ. For now, let's just assume that γ remains constant. This should be approximately true if the fee c is very small.

Note that this then implies that the mean bid ask spread θ will be multiplied by a factor $\frac{1}{(\alpha c+(1-c))} > 1$ when the fee is introduced. However, this mean is a sum of a fixed and variable component, so that some users will pay substantially less than θ while others pay substantially more. The fixed component of the spread will be multiplied by a factor $\frac{1-c}{(\alpha c+(1-c))} < 1$. That is the price quoted directly in makers' limit orders will decrease. The mean variable compenent will be $\frac{c}{(\alpha c+(1-c))}$. The variable comopnent scales with the gas fee and will differ across user types. A patient user who pays only a fraction f of the average gas price will only pay a protocol fee of $\frac{cf}{(\alpha c+(1-c))}$. If the user is sufficiently patient, he will pay less under the fee and rebate program than he does currently. In particular, users will experience a price improvement if $f < \alpha$. The amount of the price improvement for patient users will be proportional to c.

We can see here that the parameters α and c are important to the benefits the system offers for various groups of stake holders. Going forward it will be important to fine tune these parameters under the guiding hand of decentralized governance. Encouraging user ownership of ZRX will help to ensure protocol users have a voice in these decisions. Ideally, we would like to identify a happy medium where ZRX owners, retail traders, and liquidity providers all enjoy a mutual benefit.