AUTOMATED MARKET MAKERS AND DECENTRALIZED EXCHANGES: A DEFI PRIMER

VIJAY MOHAN[1]
RMIT Blockchain Innovation Hub, RMIT University
& Lattice Analytics Pty Ltd
Email: vijay.mohan@outlook.com

First version: 30th October 2020
Second version: 13th April 2021
Third version: 28th September 2021
This version: 24th November 2021

Abstract

Recent advancements in decentralized finance (DeFi) have resulted in a rapid increase in the use of Automated Market Makers (AMMs) for creating decentralized exchanges (DEXs). In this paper, we organize these developments by treating an AMM as a neoclassical black-box characterized by the conversion of inputs (tokens) to outputs (prices). The conversion is governed by the technology of the AMM summarized by an 'exchange function'. Various types of AMMs are examined, including: Constant Product Market Makers; Constant Mean Market Makers; Constant Sum Market Makers; Hybrid Function Market Makers; and, Dynamic Automated Market Makers. The paper also looks at the impact of introducing concentrated liquidity in an AMM. Overall, the framework presented here provides an intuitive geometric representation of how an AMM operates, and a clear delineation of the similarities and differences across the various types of AMMs.

Keywords Decentralized finance (DeFi); automated market maker (AMM); decentralized exchange (DEX); smart contract; Ethereum

[1] I thank four anonymous referees for their comments and suggestions for improving the content in this paper and eliminating errors. This paper is the outcome of numerous insightful discussions on DeFi over the course of several months with colleagues at RMIT's Blockchain Innovation Hub. The author would like to take the opportunity to thank the multidisciplinary cohort of academics at the Hub for this: Darcy WE Allen, Chris Berg, Sinclair Davidson, Oleksii Konashevych, Aaron M Lane, Elizabeth Morton, Kelsie Nabben, Vy Nguyen, Imon Palit, Marta Poblet, Jason Potts, Ellie Rennie, Sarah Sinclair and Stuart Thomas.

Vijay Mohan

1. Introduction

1.1 DeFi, DEXs and AMMs

The latest new thing in the blockchain space is decentralized finance (DeFi) which, broadly, refers to financial digital applications built on decentralized blockchain networks. According to one source, at the time of writing, the value of cryptocurrency locked in DeFi applications is USD 89.23 billion, having more than quadrupled over a one-year period.[2] In contrast to traditional finance that is facilitated by centralized agencies, such as banks and stock exchanges, the promise of DeFi is the elimination of centralized third-parties that act as intermediaries in financial transactions.

While it is not clear at this stage whether DeFi will replace traditional financial institutions with their decentralized substitutes, or even if many of the start-ups in this area will eventually survive, the rapidly escalating advancements in DeFi applications suggest that there is a need to step back and correlate ideas in this area with traditional concepts in economics and finance. This paper represents an attempt at examining new developments in DeFi using an old tool in economics – the neoclassical black-box. In doing so, we show that the graphical and mathematical methods familiar to generations of students of economics can provide broad insights into certain aspects of DeFi, and that many recent experiments in DeFi are applications of more general production technologies. Ultimately, the motivation here is to introduce these ideas to DeFi aficionados who are interested in an economic theory perspective, and to present some wonderful new applications of old tools to economists, especially students and educators.[3]

The application of DeFi that this paper focuses on is decentralized exchanges (DEXs). In contrast to an over-the-counter (OTC) market that enables direct trading between two agents, an exchange is an institution which standardizes assets and trading rules for multiple participants. To do so, an exchange must provide mechanisms to maintain liquidity of assets and to determine prices for assets. A stock exchange, for example, implements this through an order book system, where buyers and sellers submit 'orders': prices and volumes for an asset they would like to buy or sell. The trading price is determined by matching orders. Typically, orders are public information, allowing market participants to gauge information about interest in an asset and the price at which it is trading. In a centralized exchange, orders are maintained and implemented by a central authority (say, the New York Stock Exchange).

[2] See https://defipulse.com, accessed on 20th September, 2021.

[3] To cater to audiences from both the DeFi community and economics, by and large the paper presumes familiarity with the jargon of neither. The only assumed knowledge is some acquaintance with the concept of a blockchain and the basic techniques of calculus.

A DEX provides agents with the opportunity to exchange one asset for another without a centralized third-party responsible for overseeing trading activity. The cryptocurrency space has been dominated by centralized exchanges in the past, many of whom, such as Mt. Gox, met with disastrous downfalls and losses that have, if anything, only strengthened the resolve to make DEXs work. Schär (2021) and Pourpouneh et al (2020) list the various pros and cons of centralized and decentralized exchanges. Specifically, centralized exchanges can be easier to implement but suffer from a number or drawbacks: traders lose custody of assets and must trust the exchange to not seize assets; they can be susceptible to security threats due to a single point of attack; and, centralized exchanges for cryptocurrencies have been subject to little regulation. DEXs, on the other hand, do not rely on trust in, or security of, a single centralized party as traders retain custody of assets and smart contracts execute trades. They are, however, harder to design and implement, and can charge higher fees to attract liquidity.

DEXs can be implemented in different ways. Some replicate the order book format of a centralized exchange. One way to do this is through an on-chain order book where every order is recorded on the blockchain, but this can be expensive. An alternative approach involves constructing an off-chain order book, which only uses the blockchain for settlement, but orders are recorded elsewhere (possibly by some centralized third-party). This is less expensive, but also less decentralized and secure compared to an on-chain order book (Schär, 2021; Pourpouneh et al, 2020).

Instead of using an order book, more recent attempts at establishing DEXs have revolved around the use of automated market makers (AMMs), which is the subset that this paper focuses on. Following Hanson (2003) and Hanson (2007), AMMs first gained popularity in prediction markets for aggregating predictions of agents into prices. In general, a market maker is an institution that stands ready to buy or sell an asset, making a profit from the bid-ask spread: the difference between the ask or offer rate (the rate at which the market maker sells an asset) and the bid rate (the rate at which the market maker buys an asset). An AMM automates this by allowing traders to place orders with the AMM, which then algorithmically provides a price. It is worth reiterating how this process is distinct from the order book system that requires matches between price and volume orders provided by buyers and sellers. With a market maker, an agent trades with the market maker by selecting a quantity of an asset to trade at a price specified by the market maker. This is particularly beneficial in thin markets where there are few buyers and sellers, so that there may be a wide gap between the maximum price any buyer is willing to pay and the minimum price any seller is willing to accept, thereby causing no trades to occur. When a market maker acts as the counterparty to all trades, liquidity can be provided even when markets are thin.

Traders on a DEX are interested in swapping one token for another with the AMM. So, the question that arises is: who exactly provides liquidity for an AMM to act as a DEX? The answer is that owners of various tokens do so by placing their tokens within a liquidity pool in the

AMM; the quantities of tokens in a liquidity pool are its *reserves*.[4] In return, liquidity providers are typically entitled to a share of the fees paid by traders for exchanging tokens. Returns in the form of trading fees are the main incentive for agents to act as liquidity providers.

In an order book system, the matching of orders by buyers and sellers (that is, the forces of demand and supply) produces a price. In an AMM, prices are determined algorithmically; what guarantees, then, that the AMM prices reflect demand and supply conditions in the wider market for cryptocurrencies? For example, suppose there exists another exchange – an 'external' or 'reference' market – that prices some token, say ABC token, at 2 XYZ tokens. If a trader wants to swap between ABC and XYZ tokens at the AMM, the AMM's algorithm should set a price close to the price in the external market. There are, in fact, a couple of different ways in which the price in the AMM can be made to align with that of the external market. The first, and most common, is through the process of arbitrage, wherein arbitrageurs buy and sell assets across markets to take advantage of price differentials and make (risk-free) profits. In doing so, arbitrageurs cause an alignment of prices. The second is by allowing the reference market to act as an oracle, which is essentially an external source of information used by the AMM to set its price.

1.2 Smart contracts

Smart contracts are at the heart of DeFi applications and, consequently, current DeFi platforms are being built predominantly on Ethereum's smart contract-based blockchain.[5] Smart contracts were first introduced by Szabo (1996), along with a working definition: "A smart contract is a *set of promises*, specified in *digital form*, including *protocols* within which the *parties perform on these promises*" [emphasis added]. Closer examination of this definition reveals two distinct aspects of a smart contract: a legal contractual aspect and a technological aspect. The contractual aspect of a smart contract arises from the concepts of 'promise' and 'performance' contained in the definition. Presumably, promises are based on intention and performance will involve consideration, so one could compare this to a standard contract as defined by the legal profession. But the definition also suggests that the contract is 'digital' and includes 'protocols' (read: algorithms), which provides a distinct technological twist to the definition.

Using Szabo's (1996) definition, therefore, one can come away with different impressions of a smart contract, depending on which facet one wishes to emphasize. In this paper, we do not approach a smart contract from a legal point of view, not least because the literature on the issue of how standard contract law applies to smart contracts warrants its own separate study. Nor do we focus on issues that economists typically examine in contract theory:

[4] Of course, a single agent is free to perform both functions – provide liquidity and trade – on an AMM.

[5] See https://ethereum.org (accessed 20th September, 2021), and https://ethereum.org/en/whitepaper/ (accessed 20th September, 2021) or Buterin (2014).

asymmetric information (the complete contract literature), or property rights and organizational boundaries (the incomplete contract literature). Rather, for the purpose of this paper, the most fruitful avenue is to focus on technological aspects of a smart contract. As a definition, the US National Institute of Standards and Technology defines a smart contract as "A collection of code and data (sometimes referred to as functions and state) that is deployed using cryptographically signed transactions on the blockchain network."[6] Similarly, the Ethereum whitepaper states that, "Smart contracts, cryptographic 'boxes' that contain value and only unlock it if certain conditions are met, can also be built on top of the platform, with vastly more power than that offered by Bitcoin scripting because of the added powers of Turing-completeness, value-awareness, blockchain-awareness and state."[7]

The Ethereum network has two types of accounts: externally owned accounts (EOAs) and contract accounts (or smart contracts). EOAs in Ethereum are standard cryptocurrency accounts and are characterized by three elements: a private key, a public key and a balance of the cryptocurrency native to the blockchain, which in the case of Ethereum is Ether (ETH). The public key is used to generate an address that identifies an EOA in the network, and the private key, known only to the holder of the account, is used to (cryptographically) authorize transactions. Thus, one agent can use her private key to authorize the transfer of a certain amount of ETH to another agent on the network. When the transaction is added to a block, the blockchain state is updated and the EOAs of both agents change to reflect the transaction. In contrast, a smart contract is characterized by an address, a balance and code. A number of features follow. First, the absence of a private key associated with a smart contract implies that no agent on the network can authorize a change to the smart contract code (even if it has a bug). Rather, the smart contract is controlled by the code embedded within it, which is immutable once the contract is added to the network, and cannot be altered by any agent.[8] Second, apart from ETH, a smart contract can hold compatible tokens, such as 'ERC-20' tokens (that conform to ERC-20 standards for fungible tokens) or 'ERC-721' tokens (that meet ERC-721 standards for non-fungible tokens).[9] This is a useful feature because the smart contract

[6] See https://csrc.nist.gov/glossary/term/Smart_contract (accessed 20th September, 2021).
[7] See https://ethereum.org/en/whitepaper/ (accessed 20th September, 2021).
[8] The main hurdle to simply replacing a smart contract (say, that has a bug in it) with a new one occurs when the address generated for a smart contract depends on the current state. So, if a bug is found in a smart contract after deployment, a new contract deployed (in the future) with new code would receive a new address. Even with this constraint, there exist ways to 'upgrade' a contract, for example by using a proxy contract, or by deploying a new contract (with a new address) and migrating users and data to the new address (which can be expensive). In 2019, as part of the Constantinople upgrade to the Ethereum blockchain, the CREATE2 functionality was introduced (see https://eips.ethereum.org/EIPS/eip-1014, accessed 20th September, 2021). This essentially allows the address to be independent of future states, so it is possible to calculate a smart contract's address even without deployment. In terms of upgrading a smart contract, this implies that if a smart contract is issued an address using the CREATE2 functionality, it can be instructed to self-destruct and a new smart contract can be deployed with the same address.
[9] See https://ethereum.org/en/developers/docs/standards/tokens/erc-20/ for a description of ERC-20 and https://ethereum.org/en/developers/docs/standards/tokens/erc-721/ for ERC-721. While these are the most

acts as a sort of custodian of tokens, receiving and disbursing tokens using algorithms specified by the code.

A liquidity pool in an AMM is a smart contract with a certain set of tokens that the smart contract can maintain balances of, as specified by its code. The balances of tokens are the quantities that serve as reserves, which change as traders swap tokens in the liquidity pool. The code specifies, among other things, the rules for trading, how prices are determined based on reserves, the rules for liquidity provision, and the trading fees that traders pay to utilize the liquidity pool. An AMM itself, then, is simply a set of liquidity pools.

Finally, there are multiple transaction costs an agent may incur when utilizing an AMM. First, the agent may have to pay a *trading fee (*or *swap fee)* for exchanging tokens with the AMM; these fees act as returns for liquidity providers. Second, there is a cost to changing ledger entries on the blockchain to record changes in the ownership of tokens. These *gas fees* on Ethereum vary depending on the extent to which the blockchain is being used for any purpose, DeFi or otherwise. In what follows, we do not factor in gas fees and focus entirely on the impact of trading fees charged by a specific AMM. In reality, when performing DeFi transaction gas fees have to be included in trading decisions, and can sometimes be significant enough to deter use of DeFi applications.[10] Some AMMs also have the ability to charge *protocol fees* that are, essentially, meant for future development of the AMM. We do not examine protocol fees in this paper.[11]

1.3 The approach and contribution of this paper

This paper examines how AMMs operate as DEXs to facilitate price discovery, arbitrage, and the exchange of one cryptocurrency token for another. To achieve this, we view an AMM in much the same way as neoclassical economic theory models a firm: as a 'black-box'. A neoclassical firm is characterized entirely by its ability to convert factor inputs, such as labour and capital, into output using some technology made available by scientists and engineers. The nature (or if one were to depict this graphically, the 'shape') of the technology is summarized by a production function, which then delineates which combinations of inputs and outputs are feasible. Given the costs associated with purchasing inputs and the revenues generated by selling the output, different feasible combinations of inputs and outputs

commonly used standards at this stage, other standards exist, such as ERC-777 and ERC-1155. Standards are introduced in the Ethereum network through Ethereum Improvement Proposals (EIPs); for further details, see https://eips.ethereum.org and https://eips.ethereum.org/erc. All websites accessed 20th September, 2021.

[10] It is worth emphasizing that the omission of gas fees in our analysis is not due to their lack of importance; rather, it is due to the fact of gas fees are network-wide fees, which are present irrespective of whether AMMs exist or not. In as much as our focus is on modelling an AMM, gas fees can be viewed as an exogenous cost that has been normalized to zero in this paper.

[11] Uniswap-v2 (Adams et al, 2020), for example, introduced the possibility of a .05% protocol fee that is deducted from the trading fees. So, protocol fees do no not affect the transaction cost of the traders on the exchange; rather, they create a wedge between the fees paid by traders and the fees received by liquidity providers. In that sense, protocol fees are a tax on the users of an AMM.

generate different profits. The firm then, automaton-like, simply picks the combination of inputs and output from the set of all feasible combinations that yields the maximum profit.

In the neoclassical characterization of the firm there is scant attention paid to what goes on inside a firm (hence the term 'black-box'). Issues that real-world businesses must tackle on a daily basis, such as providing incentives to workers to exert effort, determining appropriate pay packages for employees, resolving problems associated with hierarchy and authority, and so on, are entirely ignored in a black-box view, which is based on the premise that the technology behind the production of goods and services is of paramount importance. Moreover, the black-box model does little to address what determines the boundaries of a firm – why some activities are organized within a firm and others through market transactions. Many of these problems were addressed in later developments to the theory of the firm in the form of principal-agent models and transaction cost theories.[12]

The drawbacks of the black-box methodology notwithstanding, it does what it's supposed to do – describing the conversion of inputs to outputs – extremely well. *If* in some context of analysis the focus is indeed on the technology behind this conversion, the black-box nature of the neoclassical firm is, to borrow a phrase from programmers, a feature and not a bug, because it removes all the clutter about what goes in on inside a firm to hone in on what *is* contextually important: the technology.

Given the myriad DEX platforms that have emerged in the recent past, one could make the argument that our current understanding is at a stage where it would benefit from a careful examination of how inputs are converted to an output by an AMM. The approach adopted by this paper is to view an AMM as a black-box that transforms quantities of tokens in an AMM (inputs) into a price (output). The transformation is governed by the specification of an 'exchange function' that describes the technology of token trade in an AMM.[13] It is worth emphasizing that we do not equate an AMM to a firm; rather, the methodology of treating the firm as a black-box is utilized here to model an AMM.

The exchange function that has received the most attention in the DeFi community is used by a Constant Product Market Maker (CPMM), Uniswap.[14] Since this is typically the first type of AMM that many entrants into the DeFi space encounter, Section 2 provides a detailed account, using examples, of how such an AMM operates as a DEX and how fees generate bid-ask spreads. Moreover, given the importance of arbitrage in pricing, this section examines two and three-point arbitrage, along with relevant no-arbitrage conditions.

Following its inception in 2018, Uniswap has undergone multiple iterations: Uniswap

[12] See Demsetz (1997), Hart (1995, Chapter 1), and Holmstrom and Tirole (1989) for overviews.

[13] The term 'exchange function' is used because it captures different ways in which the word 'exchange' appears in this context. First, the function governs the *exchange* of one token for another by traders in an AMM; second, any given decentralized *exchange* will have a specific function associated with it.

[14] See https://uniswap.org (accessed 20th September, 2021). As a disclaimer, the utilization of specific AMMs as examples in this paper is purely for expository purposes, and is in no way meant to advocate their use.

version 1 (v1) was the original proof of concept; [15] this was upgraded to version 2 in 2020 (v2; Adams et al, 2020).[16] While our analysis in Section 2 is based on the characteristics of Uniswap-v1 and v2, other CPMMs exist, such as Sushiswap and Mooniswap.[17] Consequently, in order to generalize the ideas here, we will often refer to a CPMM-v1 or v2 to characterize a CPMM that has the characteristics of Unsiwap-v1 or v2, respectively.

Section 3 introduces some basic tools of economic modelling - homogeneous functions, homothetic functions and Euler's theorem - as well the geometric properties characterizing these tools. While the material here is fairly self-contained, more detailed and leisurely discussions of these topics can be found in any introductory text on mathematical methods in economics, such as Chiang (1984) or Silberberg (1990).[18] This facilitates an examination of the properties and geometry of a CPMM, which complements and extends contributions to the analysis of price formation and arbitrage in a CPMM, such as Angeris et al (2019) and Zhang et al (2018).

Section 4 examines other types of AMMs that have gathered interest as DEXs, including a: Constant Mean Market Maker (CMMM), such as Balancer; Constant Sum Market Maker (CSMM), which has drawbacks as a DEX; Hybrid Function Market Makers (HFMM), such as Curve Finance; and, Dynamic Automated Market Maker (DAMM), such as Bancor.[19] The focus here is very much on fitting these AMMs into the framework developed in Section 3, as well as on drawing out similarities and points of departure in their characteristics. The specific institutional features of each of these AMMs are not described nearly as carefully as for the CPMM in Section 2, though in many instances, if required, these features can be readily related to concepts introduced for a CPMM.[20]

Overall, while a variety of AMMs have emerged to trade tokens in a decentralized manner, their performance and attributes are connected. Figure 1 below serves to summarize the links we develop between the various AMMs in Sections 3 and 4. As the figure shows, a CPMM is a special case of a CMMM. Allowing for weights to change in a CMMM replicates the performance (in terms of price adjustments) of a DAMM. Mixing a CSMM with a CMMM yields a HFMM.

---

[15] See https://hackmd.io/@HaydenAdams/HJ9jLsfTz for the v1 whitepaper (accessed 20th September, 2021).

[16] As each version introduces new smart contracts, they can coexist as liquidity pools.

[17] See https://sushi.com and https://mooniswap.exchange/#/swap. Sushiswap is a fork of Uniswap, with some features added to make it more "community friendly". In what was subsequently termed a "vampire attack", Sushiswap sought to extract liquidity away from Unsiwap. For a summary of vampire attacks, see https://finematics.com/vampire-attack-sushiswap-explained/. Mooniswap, on the other hand, attempts to reduce the impact of price slippage from large trades on a CPMM. All websites accessed 20th September, 2021.

[18] An application of these techniques can also be found in a graduate microeconomics text, such as Mas-Colell, Whinston and Green (1995).

[19] See: https://balancer.finance for Balancer and https://www.curve.fi for Curve Finance; and https://app.bancor.network/eth/data for Bancor. All websites accessed 20th September, 2021.

[20] See also Angeris and Chitra (2020), who examine the properties of a general class of constant function market makers, which include the CPMM, CMMM and HFMM.

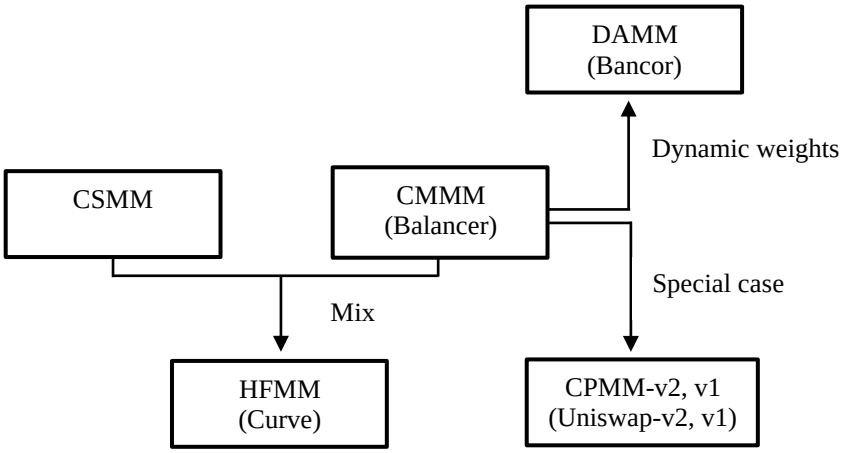

Figure 1: Links between AMMs

Recently, Uniswap released Version 3 (v3; Adams et al, 2021), which is a significant departure from v2. In Section 5, we review the changes brought about by Uniswap-v3 and turn to broadly outlining the geometry behind this AMM by examining the idea of concentrated liquidity. We show here that Uniswap-v3 displays some characteristics of both a CPMM and a CSMM.

Like the neoclassical firm, the black-box view of an AMM has limitations in describing other aspects of an AMM, such as how decentralized governance of an AMM is organized, how communities are formed in this space, and so on; this paper does not address these important and interesting issues. Nor does this paper delve into the uses of AMMs other than for the purpose of creating a DEX. Some concluding thoughts on this are offered in Section 6.

2. The workings of a Constant Product Market Maker (CPMM): Uniswap-v1 and v2

In a CPMM such as Uniswap, the product of the quantity of two tokens in a liquidity pool is a constant. The example examined here replicates and extends the one developed in the original Uniswap-v1 whitepaper. When needed, we will introduce the changes brought about in Uniswap-v2, and how this impacts the workings of the CPMM platform.

2.1 Examples and some basic pricing equations

The two tokens in focus here are token X (which we assume to be ETH) and token Y (some ERC-20 token, say ABC). In what follows, we use upper-case letters for the token in general and lower-case letters for specific quantities of the tokens. For any given token quantities, (x, y), a CPMM utilizes the exchange function $xy = k$ to algorithmically govern trade between the two tokens. To see how this works, let the initial amount of ETH supplied by liquidity providers be $x^0 = 10$, and the amount of ABC provided be $y^0 = 500$; these are the initial reserves of the two tokens in the AMM liquidity pool. In a CPMM, the product of the

two is a constant or *invariant*, k, which takes an initial value $k^0 = x^0 y^0 = 5000$.

A trader in this market is any agent who exchanges one token for anther in the pool. In order to do this, the trader must pay a trading fee, which accrues to the liquidity providers. However, to establish a baseline case, let us first consider the situation where the trading fee is zero. Suppose the trader wishes to sell 1 ETH in exchange for ABC. In the absence of a trading fee, this entire amount is added to the liquidity pool, resulting in a pool balance of $x^1 = 11$ ETH. Given k^0, the amount of ABC in the pool is $y^1 = \frac{k^0}{x^1} = 454.5454$. The change in the pool reserves of ABC is $y^1 - y^0 = -45.4546$, which is negative because the reserves of ABC in the AMM have fallen; this quantity of 45.4546 ABC is sent to the trader's account. The trader has effectively sold 1 ETH in order to buy 45.4546 ABC with the AMM, which is the counterparty to the trade; consequently, the liquidity providers for the AMM now, in aggregate, hold 1 ETH more and 45.4546 ABC less. While the AMM has an altered reserve profile (x^1, y^1), it is evident that $x^1 y^1 = k^0$, and the product of the new reserve quantities equals the constant k^0. The price of 1 ETH in terms of ABC in this example is $P^b_{Y/X} = 45.4546$ ABC/ETH. In terms of the notation, the subscript Y/X indicates that the price is for one unit of token X (ETH) in terms of token Y (ABC), while the superscript b indicates that this is the bid price: the price at which the AMM buys ETH.[21] It is straightforward to work the example in reverse, where the trader buys 1 ETH from the AMM. Proceeding as before: $x^1 = 9$, $y^1 = \frac{k^0}{x^1} = 555.5556$ and $y^1 - y^0 = 55.5556$. The price in this instance is $P^a_{Y/X} = 55.5566$ ABC/ETH, where the superscript a indicates that this is the ask (or offer) price of ETH: the price at which the AMM sells 1 unit of ETH to the trader.

The bid and ask rates in this example seem appreciably different – the bid-ask spread is $P^a_{Y/X} - P^b_{Y/X} = 10.102$, primarily because the changes considered here are lumpy and fairly large in magnitude. In reality, ETH is divisible to 18 decimal places, so much smaller trades are feasible.[22] Suppose the trader were to buy and sell .01 ETH instead of 1 ETH in the example. The bid and ask prices are now $P^b_{Y/X} = 49.9500$ ABC/ETH and $P^a_{Y/X} = 50.0501$ ABC/ETH and the spread is only $P^a_{Y/X} - P^b_{Y/X} = 0.1001$. Indeed, as the change in ETH in the trade keeps getting smaller, the two converge towards the price of 50 ABC/ETH. In general, it can be shown (which we do subsequently) that with infinitesimal changes and no trading fees:

(1) $$P^b_{Y/X} = P^a_{Y/X} = \frac{y^0}{x^0} = P_{Y/X}$$

There are a few features to take away from this example. First, equation (1) summarizes the fact that the bid and ask prices collapse to a single price ($P_{Y/X}$) for infinitesimally small

[21] In the notation Y/X, X is the 'base' token and Y is the 'pricing' token. As in traditional finance, we use the terms 'bid' and 'ask' (or 'offer') as rates viewed from the perspective of the market maker. A trader, therefore, sells at the bid price, and buys at the ask price.

[22] See https://ethereum.org/en/eth/ (accessed 20th September, 2021).

changes, and that this price equals the ratio of the two reserves ($\frac{y^0}{x^0}$). Second, the price of one unit of ABC in terms of ETH (that is, $P_{X/Y}$) converges to:

(2) $$P_{X/Y} = \frac{1}{P_{Y/X}} = \frac{x^0}{y^0}$$

Third, with no fees, k^0 remains the same for all subsequent trades, and only changes when liquidity is injected into, or withdrawn from, the pool by liquidity providers.

Now consider a situation where the trader is required to pay an *ad valorem* trading fee, τ. Apart from the addition of fees, all the other features of the example are maintained. The trading fee is assumed to be $\tau = .25\%$ in the example constructed in the Uniswap-v1 whitepaper, which we follow here as well; in reality, the fee on Uniswap-v1 and v2 is 0.3%. The trading fee is paid using the token that is added to the liquidity pool (that is, the token sold by the trader). Assuming, as before, that the trader sells 1 ETH, the transaction cost of a sale of 1 ETH by the trader is $\tau \times 1 = .0025$ ETH. This leaves $(1 - \tau) \times 1 = .9975$ ETH available for the trade. To save on notation, let $(1 - \tau) = \phi$. There are two stages to the process now: in the first stage, the fee is deducted, and the trade is enforced using the initial invariant k^0. In the second stage, the fees are added to the liquidity pool, which gives rise to a new invariant, k^1. This ensures that the value of k changes with every trade; so, the term 'constant' or 'invariant' is somewhat misleading when there are trading fees – it is more of a predetermined variable prior to a trade.

In the first stage, which we can think of as the interim trading stage, the amount of ETH in the pool is $x^{1\prime} = 10.9975$ ETH (where the superscript $'$ denotes interim). Given $k^0 = 5000$, the new amount of ABC balance in the pool is $y^1 = \frac{k^0}{x^{1\prime}} = 454.6488$, implying a change $y^1 - y^0 = -45.3512$ ABC. The trader effectively receives the bid price of $p^b_{Y/X} = 45.3512$ ABC/ETH. In terms of notation, the lower-case p indicates that the price now refers to a scenario where a trading fee has been imposed. In the second stage of the trade, the fees are added to the liquidity pool, giving a final ETH pool balance of $x^1 = 11$. The new value of k at the end of the trade is $k^1 = x^1 y^1 = 5001.1368$, which is higher than before and is a predetermined variable for the next trade on the AMM.

Comparing the case with trading fees to the case without, we can see that the bid price of ETH is lower than before ($p^b_{Y/X} < P^b_{Y/X}$), and the trader receives less ABC per ETH than when there are no transaction costs. Secondly, the change in k is $k^1 - k^0 = 1.1368$. Greater the number of transactions on Unsiwap-v1 or v2, more rapid is the expansion of k, even if liquidity providers add no further tokens to the pool themselves.

Now suppose the trader were to buy one unit of ETH, which would leave 9 ETH in the pool. The fee here is charged for the volume of ABC traded for 1 ETH. So, if the trader deposits an amount Δy ABC, after the fees are deducted the trader adds an interim (Stage 1) amount of

$\phi\Delta y$ ABC to the pool. It follows that $\Delta y = \frac{1}{\phi}\left[\frac{k^0}{x^1} - y^0\right] = 55.6948$. The ask price has now increased with the introduction of the transaction fee ($p^a_{Y/X} > P^a_{Y/X}$). Moreover, the bid-offer spread is $p^a_{Y/X} - p^b_{Y/X} = 10.3436$, which is larger than the case with no trading fees, because the bid rate is lower and the ask rate is higher. This is intuitively the reason why there are lower arbitrage opportunities available in the presence of higher transaction costs.

We can now ask, once again, what happens if a smaller ETH amount, say .01 ETH, is traded. Replicating the procedure, we find that $p^b_{Y/X} = 48.7025$, while $p^a_{Y/X} = 50.1755$, so $p^a_{Y/X} - p^b_{Y/X} = 1.473$. While this is smaller than the spread when 1 ETH was traded, it is still fairly significant. In fact, we show analytically in the Section 3 that for infinitesimal changes, $p^b_{Y/X} = 49.875$ and $p^a_{Y/X} = 50.125$, which implies $p^a_{Y/X} - p^b_{Y/X} = 0.25$ is the minimum spread achievable in this example with a fee of 0.25%. This wedge between the two prices cannot reduce to zero – we need to work with bid and ask rates even with infinitesimal changes.

Finally, since the sale (purchase) of one token by a trader goes hand-in-hand with the purchase (sale) of the other token, equation (2) needs to be modified to:

(3) $\quad p^b_{Y/X} = \frac{1}{p^a_{X/Y}} \quad$ and $\quad p^a_{Y/X} = \frac{1}{p^b_{X/Y}}$

2.2 Arbitrage using AMMs

In this section, we consider two different ways in which an AMM can be used for arbitrage: two-point arbitrage and three-point arbitrage. The arbitrage terminology we employ here is consistent with standard usage in the finance and foreign exchange market literatures.[23]

2.2.1 Two-point arbitrage

Two-point arbitrage or *locational arbitrage* is triggered due to a difference in prices across different markets (or exchanges or platforms, as the case may be) for the same asset. Consider a situation where there exists another (possibly centralized) exchange, which is the external reference market. An agent seeks to exchange between Y (ABC) and X (ETH). Two-point arbitrage refers to the fact that a token bought in one market can be sold in the other in order to realize risk-free profits. Doing so is profitable (subject to transaction costs) if there is a mismatch in the prices quoted in the two markets. However, as agents take advantage of arbitrage opportunities and transact in the two markets, the demand and supply forces they set into motion eventually cause these opportunities to disappear. An equilibrium occurs when there are no arbitrage opportunities left, and the equilibrium *no-arbitrage condition*

[23] Two-point and three-point arbitrage in the context of foreign exchange markets are covered in most undergraduate texts in international finance; see, for example, Moosa (2010). Our discussion draws on these insights, along with simple ways to visualize the arbitrage process, such as Figure 2.

essentially involves no mismatch in prices between the two markets. Arbitrage is, in general, a powerful pricing tool: first, it is attractive because it results in risk-free profits, so we can be sure agents will jump to take advantage of arbitrage opportunities when they arise; second, the process of arbitrage conveniently eliminates these opportunities eventually, so we can be confident that, all else being the same, markets will gravitate towards an equilibrium where prices equalize (to the extent permitted by transaction costs).

To examine two-point arbitrage in more detail, let the price of one unit of X in terms of Y in the external market be $M_{Y/X}$. The upper-case M denotes a situation where there are no trading fees for transacting in the external market. This is altered readily enough by considering bid and ask rates in the external market: $m^b_{Y/X}$ and $m^a_{Y/X}$. We could make the story even more realistic by adding other transaction costs (such as gas fees). In reality, all such costs must be accounted for. However, for the task at hand of understanding the mechanics of how arbitrage works, we can ignore transaction costs other than the AMM trading fees, with the caveat that they must be factored in when actually trading.

To begin with, consider the case when there are no trading fees on the AMM and infinitesimal trade volumes are possible. Suppose that $M_{Y/X} = 49.9$ in the external market, while $P_{Y/X} = 50.0$ in the AMM. An arbitrageur could buy X in the external market for 49.9 and sell it in the AMM for $50.0\ Y$, thereby making a profit of $P_{Y/X} - M_{Y/X} = 50.0 - 49.9 = 0.1$ units of Y per unit of X. However, as arbitrageurs do this, the increased demand for X in the external market will increase $M_{Y/X}$, and the increased sale of X in the AMM will reduce $P_{Y/X}$ till the arbitrage profits are wiped away.[24] The prices in the two markets are brought into alignment due to the actions of arbitrageurs, and prices in the AMM cannot diverge randomly for any significant period of time. The equilibrium no-arbitrage condition satisfies:

(4) $$P_{Y/X} = M_{Y/X}$$

Similarly, in instances where $P_{Y/X} < M_{Y/X}$ we would expect arbitrage to occur in the opposite direction. Demand-supply forces would then result in (4) holding at equilibrium.

Now, suppose there is a trading fee of $\tau = 0.25\%$ on AMM which, as we have seen earlier, results in $p^b_{Y/X} = 49.875$ and $p^a_{Y/X} = 50.125$. With $M_{Y/X} = 49.9$, arbitrage is no longer feasible: buying ETH in the external market and selling in the AMM involves a loss of 0.025 ABC, and going in the reverse direction yields a loss of .225 ABC. Thus, the no-arbitrage condition in equation (4) is transformed to:

(5) $$p^b_{Y/X} \leq M_{Y/X} \leq p^a_{Y/X}$$

In the range prescribed by equation (5), arbitrage between the external market and the AMM is not profitable. If $M_{Y/X}$ moves outside this range, arbitrage is triggered which will result in

[24] If the AMM is small relative to the external market, these trades may have only a small impact on the price in the external market, so that the brunt of the adjustment occurs through changes in $P_{Y/X}$.

demand-supply changes that bring the rates into alignment, so that (5) holds again.

Two observations are worth noting about the arbitrage process described here. First, the numbers we have used for $p^b_{Y/X}$ and $p^a_{Y/X}$ are taken from example in Section 2.1 with infinitesimal trade volume. However, moving to discrete changes only serves to widen the spread $p^a_{Y/X} - p^b_{Y/X}$, which extends the range in equation (5) for which arbitrage is not profitable. A similar widening of the spread occurs when τ rises, so that increased trading fees (and transaction costs in general) also reduce arbitrage opportunities. Second, if one incorporates bid and offer rates in the external market, $m^b_{Y/X}$ and $m^a_{Y/X}$, the equilibrium no-arbitrage condition in (5) will have to factor this in; specifically:

(6) $\quad p^b_{Y/X} \leq m^a_{Y/X}$ and $m^b_{Y/X} \leq p^a_{Y/X}$

2.2.2 Three-point (or triangular) arbitrage

Two-point arbitrage exploits differences in prices across markets and, as the name suggests, requires two price quotes – one from each market – for arbitrage possibilities to emerge. In contrast, three-point arbitrage (or triangular arbitrage) focuses on the internal consistency in the prices offered within a single market and requires three price quotes for its implementation. Although three-point arbitrage can also be performed across markets, it is not necessary – a price misalignment in a single market can trigger three-point arbitrage opportunities. We will see that as agents take advantage of three-point arbitrage within a single AMM, the changing forces of demand and supply ensure that prices in different liquidity pools in the AMM are ultimately aligned in a logically consistent way.

Suppose there are three tokens, X, Y and Z, that are offered for trade and that a trader can swap between any pair. The question is: can the trader start off with 1 unit of any arbitrary token (say Z) and cycle through the tokens by, for example, selling Z for X, then selling the X proceeds for Y before converting back to Z in order to end up (magically) with more than 1 unit of Z? In terms of notation, we use the symbols $Z \to X \to Y \to Z$ to describe this sequence of conversions. However, this is not the only sequence that is feasible: the trader could also perform the sequence $Z \to Y \to X \to Z$. In either one of these cases, if the trader starts with 1 unit of Z at the beginning of the cycle in order to end up with more than 1 unit of Z at the end of the cycle, the possibility of three-point arbitrage exists. Moreover, the choice of Z as the starting token is arbitrary: a trader could realize arbitrage profits by cycling through starting with X or Y. As Figure 2 shows, all these possibilities can be visualized using a triangle (hence the name triangular arbitrage), with the three tokens located on the three vertices, and a specific cycle forming a path along the sides starting an ending at a given vertex.

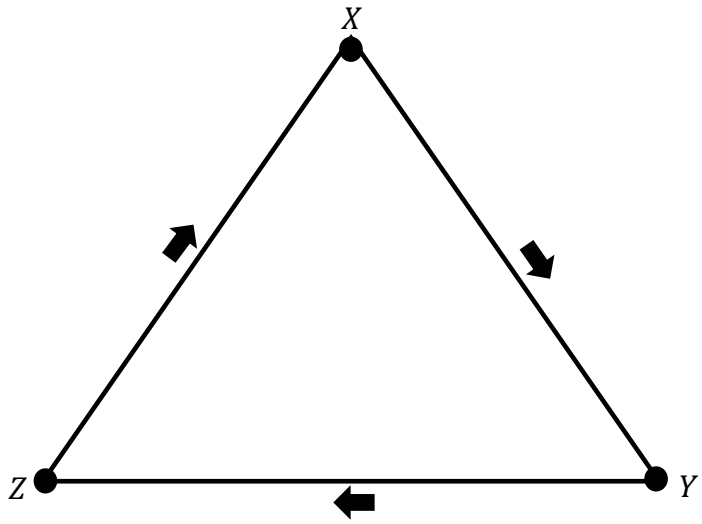

Figure 2: Three-point arbitrage

The arrows in Figure 2 show the cycles $Z \to X \to Y \to Z$, $Y \to Z \to X \to Y$ and $X \to Y \to Z \to X$, depending on which vertex forms the starting point. The opposite cycles can be visualized by reversing the direction of the arrows.

Uniswap runs on the Ethereum network and ETH (X) is the native settlement token. However, there are a number of ERC-20 tokens that can be traded on the network, and one can swap between any two ERC-20 tokens Y and Z on Uniswap. How exactly this exchange is facilitated varies between v1 and v2 of Uniswap. In Uniswap-v1, the platform performs an exchange between Y and Z by implicitly using ETH as a via medium. Uniswap achieves this by maintaining a separate smart contract for the exchange of each ERC-20 token with ETH. Uniswap-v1 does *not* incorporate the possibility of a smart contract for direct conversions between Y and Z. This feature changes in Uniswap-v2, where a distinct smart contract can be created for direct exchanges between ERC-20 tokens. Adams et al (2020) recognize the added complexity of this change,[25] without explicitly outlining the new possibilities for arbitrage this generates. Much of the focus of this section, then, is to outline how Uniswap-v2 creates the possibility of three-point arbitrage that is absent in v1.

More generally, we explore the outcomes under two types of CPMMs: CPMM-v1 that uses a token X as an intermediary in the exchange between and Y and Z; and CPMM-v2 that allows for a smart contract as a $Y - Z$ liquidity pool. Consider CPMM-v1 to begin with. If a trader wishes to sell Z and purchase Y, the AMM will proceed by converting Z to X (ETH) and then by converting the proceeds to Y, which are added to the trader's account. Suppose that there are no trading fees ($\tau = 0$). If the trader sells 1 unit of token Z, $P_{X/Z}$ determines how many

[25] As Adams et al (2020) state: "A proliferation of pairs between arbitrary ERC-20s could make it somewhat more difficult to find the best path to trade a particular pair, but routing can be handled at a higher layer…".

units of X the AMM delivers in exchange using the $X - Z$ liquidity pool. The $P_{X/Z}$ units of X can then be converted to Y at the price $P_{Y/X}$, yielding the trader $P_{Y/X} \times P_{X/Z}$ units of Y. Consequently, the price offered by CPMM-v1 for converting 1 unit of Z to Y, $P_{Y/Z}$, is:

(7) $$P_{Y/Z} = P_{Y/X} \times P_{X/Z}$$

To borrow some terminology from the foreign exchange rate literature, $P_{Y/Z}$ is the *cross-rate* (or *cross-price*) that can be derived from the other two prices ($P_{Y/X}$ and $P_{X/Z}$). In CPMM-v1, there are two prices, $P_{Y/X}$ and $P_{X/Z}$, that are determined in the two liquidity pools that are available, the $X - Y$ pool and the $X - Z$ pool. The third price, $P_{Y/Z}$, is derived from the other two (using equation 7). Now, in CPMM-v2, a third market price exists due to the creation of a direct $Y - Z$ liquidity pool. The three liquidity pools (the $X - Y$, $X - Z$ and $Y - Z$ pools) can move independently due to trader activities. Intuitively, we can see from Figure 2 that the presence of a third liquidity pool creates an independent price to connect the vertices Y and Z.

When equation (7) holds, three-point arbitrage is not feasible; it is, therefore, a no-arbitrage condition. To see this, consider CPMM-v2 with $X - Y$, $X - Z$ and $Y - Z$ liquidity pools, along with prices ($P_{Y/X}$, $P_{X/Z}$ and $P_{Y/Z}$). Let us further posit that these three pools yield the relationship $P_{Y/Z} < P_{Y/X} \times P_{X/Z}$, which violates equation (7). A trader could then perform the sequence $Z \to X \to Y \to Z$ to make a risk-free profit.[26] Starting off with 1 unit of Z, this sequence of conversions yields $P_{Z/Y} \times P_{Y/X} \times P_{X/Z}$ units of Z at the end of the cycle. As $P_{Z/Y} = \frac{1}{P_{Y/Z}}$ (equation 2), we can rewrite the expression $P_{Z/Y} \times P_{Y/X} \times P_{X/Z}$ as $\frac{P_{Y/X} \times P_{X/Z}}{P_{Y/Z}}$ and, given the assumption that $P_{Y/Z} < P_{Y/X} \times P_{X/Z}$, it follows that $\frac{P_{Y/X} \times P_{X/Z}}{P_{Y/Z}} > 1$.[27]

As enough traders do this, relative prices will change until the arbitrage opportunity is wiped out, which occurs when $P_{Y/Z} = P_{Y/X} \times P_{X/Z}$ (in other words, $P_{Z/Y} \times P_{Y/X} \times P_{X/Z} = 1$) and the inequality no longer holds. A simple demand-supply argument suffices to verify this. In the first step of the arbitrage sequence $Z \to X \to Y \to Z$, the trader sells Z in exchange for X, which decreases $P_{X/Z}$. In the next step, the trader sells X in exchange for Y, which reduces $P_{Y/X}$. In the last step, the trader sells Y in exchange for Z, which reduces $P_{Z/Y}$. So, starting with the inequality $P_{Y/Z} < P_{Y/X} \times P_{X/Z}$, the arbitrage process increases the term on the left-hand side of the inequality and reduces the terms on the right; the process stops only when equality is restored and equation (7) holds, thereby making it a no-arbitrage condition.

In CPMM-v1, equation (7) always holds by construction, so one can think of this being an *identity* in this type of AMM. On the other hand, in CPMM-v2, equation (7) holds as an

[26] What is important here is not the vertex in Figure 2 that we start off with, but the direction around the triangle that we take. Thus, the cycles $Y \to Z \to X \to Y$ and $X \to Y \to Z \to X$ also yield arbitrage profits in this instance.
[27] If the reverse inequality were to hold, the trader could make risk-free profits by cycling tokens in the opposite direction: $Z \to Y \to X \to Z$.

equilibrium condition after arbitrage opportunities have been eliminated. To put this slightly differently, there exists only one price for exchanging between Y and Z a trader can obtain in CPMM-v1, which is automatically determined by the AMM using equation (7). In CPMM-v2, there are two prices that are available to a trader: $P_{Y/Z}$, that comes about by trading in the $Y - Z$ pool, and $P_{Y/X} \times P_{X/Z}$, that can be obtained by performing the sequence $Z \to X \to Y$; the two are necessarily equal only at equilibrium. If there is disequilibrium, the trader can make arbitrage profits by buying through the method that is cheaper and selling through other method. Thus, when $P_{Y/Z} < P_{Y/X} \times P_{X/Z}$, the trader can buy Z through the $Y - Z$ pool and sell it through the sequence $Z \to X \to Y$ to make a profit.

Adding a fee, τ, does not alter the fundamental nature of the process; it does make the calculations more cumbersome though. To see the implications of this, consider CPMM-v2. The $Y - Z$ pool has a bid-ask spread $p^a_{Y/Z} - p^b_{Y/Z}$, and the trader can directly buy Z at the ask-price and sell it at the bid-price. However, the trader also has the alternative option of performing the sequence $Z \to X \to Y$ to sell Z and buy Y, and the sequence $Y \to X \to Z$ to sell Y and buy Z. This alternative method requires two trading fees. For the sequence $Z \to X \to Y$, for example, there exists a fee for each step $Z \to X$ and $X \to Y$; consequently, the trader sells Z at price $p^b_{X/Z}$, and then sells the X proceeds at $p^b_{Y/X}$. This implies that the exchange rate associated with the sequence $Z \to X \to Y$ is $p^b_{Y/X} \times p^b_{X/Z}$. Similarly, the rate associated with the sequence $Y \to X \to Z$ is $p^a_{Y/X} \times p^a_{X/Z}$. This yields no-arbitrage conditions:

(8) $\quad p^a_{Y/Z} \geq p^b_{Y/X} \times p^b_{X/Z} \quad$ and $\quad p^a_{Y/X} \times p^a_{X/Z} \geq p^b_{Y/Z}$

Equation (8) essentially states that it is never profitable to buy Z through one method available to the trader and sell it through the other.[28]

2.3 Liquidity provision and fees in a CPMM: Uniswap-v1 and v2

Any liquidity provider on Uniswap-v1 and v2 simultaneously adds both tokens in a liquidity pool. This is a desirable feature which ensures that prices do not fluctuate due to liquidity provision. Suppose the price of X in terms of Y (when $\tau = 0$) is $P_{Y/X} = \frac{y}{x}$. Consider a situation where an agent wished to add liquidity to the $X - Y$ pool, and did so by only adding a certain amount of X, equal to Δx. This would result in a price of $P'_{Y/X} = \frac{y'}{x'} = \frac{y}{x+\Delta x} < P_{Y/X}$. Adding a significant amount of X to the pool by liquidity providers, then, would depress the price of X,

[28] If the trader wishes to buy Z using the $Y - Z$ pool, the trader incurs a price of $p^a_{Y/Z}$; if the trader the sells the proceeds using the sequence $Z \to X \to Y$, the trader receives $p^b_{Y/X} \times p^b_{X/Z}$. When $p^a_{y/z} \geq p^b_{y/x} \times p^b_{x/z}$ holds, this strategy is not profitable. Similarly, if the trader buys Z using the sequence $Y \to X \to Z$ the effective price is $p^a_{Y/X} \times p^a_{X/Z}$. When $p^a_{Y/X} \times p^a_{X/Z} \geq p^b_{Y/Z}$ holds, selling the proceeds in the $Y - Z$ pool is unprofitable. Equation (8), therefore, describes the equilibrium no-arbitrage condition in CPMM-v2. No such arbitrage opportunities can exist in CPMM-v1, since the trader has only the bid-ask spread $p^a_{Y/X} \times p^a_{X/Z} - p^b_{Y/X} \times p^b_{X/Z}$ to work with.

which is undesirable; ideally, we would like the price to change because of trading activity and not due to the act of providing liquidity. To prevent a price change from occurring, we would require $y' = x' \times P_{Y/X}$. If Δy is the amount of Y that needs to be added to make this condition hold, some algebraic manipulation then yields the fact that we would need $\frac{\Delta y}{\Delta x} = \frac{y}{x}$; that is, the two tokens have to be added in the same proportion as the existing reserves so as to not cause a price change.[29] For infinitesimally small changes, this can be rewritten as differentials, $\frac{dy}{y} = \frac{dx}{x}$.

Liquidity providers are entitled to the fees that are paid by traders on Uniswap, in proportion to the amount a liquidity provider has contributed. To facilitate this, and the process of adding and removing liquidity, Uniswap-v1 and v2 mints and distributes *liquidity tokens*, with a specific token issued to each liquidity pool. These tokens are themselves tradeable as fungible ERC-20 tokens.

3. The geometry of a CPMM

3.1 Preliminaries: homogenous and homothetic functions

The function $xy = k$ used in a CPMM is an example of a class of functions referred to as *homogenous functions*, which we define formally below.

Homogenous functions: A function $f(x_1, x_2, \dots, x_n)$ is defined to be homogenous of degree r when the following condition is satisfied:

$$f(\lambda x_1, \lambda x_2, \dots, \lambda x_n) = \lambda^r f(x_1, x_2, \dots, x_n)$$

Essentially, this states that if all the independent variables $(x_1, x_2, \dots, x_n)$ are multiplied by some factor λ, the value of the function is multiplied by λ^r. The CPMM function xy is a simplified version of this, with two independent variables. Utilizing Definition 1, it is evident that this function is homogenous of degree 2: if $xy = k$, then $(\lambda x)(\lambda y) = \lambda^2 xy = \lambda^2 k$. So, for instance, if one were to double both x and y, k would quadruple.

A homogenous function $k = f(x_1, x_2, \dots, x_n)$ can be used in different contexts to characterize different things. In consumer theory it describes, for example, the utility function: the level of utility (k) derived from the consumption of certain quantities of commodities $(x_1, x_2, \dots, x_n)$. In production theory it describes the production function: the quantity of good or service (k) produced by a firm from a certain combination of factor inputs $(x_1, x_2, \dots, x_n)$. In the current context of an AMM, we think of it as an *exchange function* that links various amounts of tokens $(x_1, x_2, \dots, x_n)$ to the AMM's invariant, k.

Geometrically, the exchange function $xy = k$ generates level or contour curves that are

[29] Specifically, $(y + \Delta y) = (x + \Delta x) \times P_{Y/X}$, which implies $(y + \Delta y) = (x + \Delta x) \times \frac{y}{x}$, which in turn implies that $(y + \Delta y)x = (x + \Delta x)y$. Simplifying yields the desired result.

rectangular hyperbolas in a two-dimensional graph, as shown in Figure 3 below. A level curve shows, in the $x - y$ plane, various combinations of x and y that yield a specific value of k. Figure 3 depicts two level curves; the first shows all the combinations of x and y that yield a specific value of k^0; the second shows all combinations of x and y that yield k^1. Since k can take any value (greater than zero), there are an infinite such level curves that can be drawn, each corresponding to a particular value of k.[30] Subsequently, in Section 3.4, we will argue that the level curves of the exchange function can best be thought of as 'isoliquidity' curves.

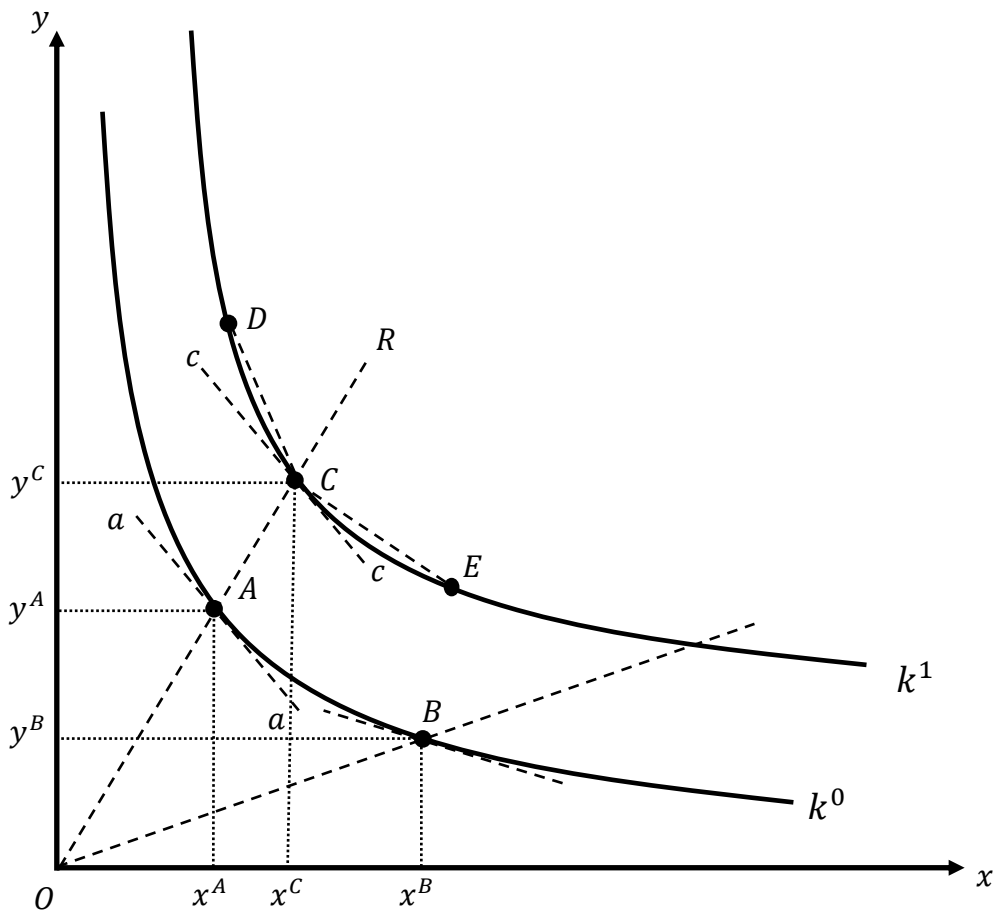

Figure 3: The $xy = k$ function

In general, suppose we allow x and y to change. Taking the total derivative of $xy = k$, we get that $ydx + xdy = dk$. Along any particular level curve, however, $dk = 0$, and so:

(9) $$\frac{dy}{dx} = -\frac{y}{x}$$

The intuition behind equation (9) is as follows. Suppose we consider a particular point, say point A in Figure 3, which has pool reserves of x^A and y^A. Starting from point A, if x changes

[30] The corresponding level curves in the context of consumption are indifference curves, and in the context of production are isoquants.

infinitesimally, y has to change by a certain amount so that $xy = k^0$ continues to hold and we remain on the level curve. The left-hand side of equation (9) is the slope of the curve at A, which corresponds to the slope of the tangent aa in Figure 3. This slope is negative due the fact that if x increases, y must decrease in order for the value of k not to change. The right-hand side of equation (9) is the slope of the ray OA connecting A to the origin O. So, equation (9) states that if x and y change along curve k^0, the slope of the tangent at A equals the negative of the slope of the ray OA; that is: $\frac{dy}{dx}\Big|^A = -\frac{y^A}{x^A}$. Similarly, at point B, $\frac{dy}{dx}\Big|^B = -\frac{y^B}{x^B}$ holds and the absolute value of the slope is lower at point B compared to A.

One of the features of a homogenous function is that the slope remains unchanged as we move from one level curve to another along a given ray. While we do not provide a formal proof of this here for a general homogenous function, it can be shown for our context in a straightforward manner using Figure 3.[31] Specifically, since points A and C both lie on the ray OR, the slope of OA is equal to the slope of OC; that is, $\frac{y^A}{x^A} = \frac{y^C}{x^C}$. From (9) it then follows that $\frac{dy}{dx}\Big|^A = \frac{dy}{dx}\Big|^C$, and the slopes of tangents are equal at A and C (slope of aa equals cc). In fact, this property of slopes of level curves being invariant along a ray from the origin is termed *homotheticity* and is true of a broader class of functions: *homothetic functions.*

Homothetic functions: If $k = f(x_1, x_2, \dots, x_n)$ is a homogenous function of any degree r, a homothetic function is a composite function $H = h(k) = h(f(x_1, x_2, \dots, x_n))$, such that $h'(k) \neq 0$ for any k.

Here $h'(k)$ is the derivative with respect to k. A homothetic function is a monotonic transformation of a homogenous function. While the transformation can, in general, be positive or negative, it is often useful to focus on the positive transformation where $h'(k) > 0$, because this has the desirable property that a higher k is associated with a higher H.

Though a homothetic function transforms a homogenous function, it is itself not necessarily homogenous. The classic example of this is a log transformation. If we start with $k = xy$ (which is homogeneous of degree 2), $H(x, y) = \log(k) = \log(xy)$ is a homothetic function since $\frac{d \log k}{dk} > 0$. However, H is not a homogenous function, because $\log(\lambda x \times \lambda y) \neq \lambda^2 \log(xy)$. Nevertheless, H inherits the homotheticity property. This is because at any arbitrary point the level curves of $H(\cdot)$ function have exactly the same slope as the corresponding level curve of $f(\cdot)$ through that point. For example, consider the two-independent variable case, where $k = f(x, y)$ is homogenous of degree r, and $H(x, y) = h(k) = h(f(x, y))$. The slope of the level curve of $H(\cdot)$ at any point is $\frac{dy}{dx}\Big|_H = -\frac{\partial H/\partial x}{\partial H/\partial y}$, while

[31] The proof is standard in any text on introductory mathematical economics (for example, Silberberg, 1990).

the slope of the level curve of the $f(\cdot)$ function through that point is $\left.\frac{dy}{dx}\right|_f = -\frac{\partial f/\partial x}{\partial f/\partial y}$. To show that the slopes of the two level curves are the same:

(10) $$\left.\frac{dy}{dx}\right|_H = -\frac{\partial H/\partial x}{\partial H/\partial y} = -\frac{h'(k)}{h'(k)}\frac{\partial f/\partial x}{\partial f/\partial y} = -\frac{\partial f/\partial x}{\partial f/\partial y} = \left.\frac{dy}{dx}\right|_f$$

The homotheticity property is useful because it tells us what happens to level curves if we start with a CPMM with an exchange function $xy = k$ and transform it monotonically.

Another handy property of homogenous functions is *Euler's theorem*, which we will find useful in the next section for valuing a liquidity pool:[32]

Euler's Theorem: Suppose $f(x_1, x_2, \dots, x_n)$ is a homogenous function of degree r. Then:

$$\frac{\partial f}{\partial x_1}x_1 + \frac{\partial f}{\partial x_2}x_2 + \cdots \frac{\partial f}{\partial x_n}x_n = rf(x_1, x_2, \dots, x_n)$$

3.2 Arbitrage and price determination in a CPMM

We are now in a position to provide a simple graphical representation of much our discussion on arbitrage and price determination. To do this, we can contextualize Figure 3 as the exchange function of a CPMM. As before, we begin with the situation where $\tau = 0$.

The price of X in terms of Y is $P_{Y/X}$; graphically, for infinitesimal changes, this translates to the absolute value of the slope $(-\frac{dy}{dx})$. For any combination of reserves (x, y), equation (9) indicates that $P_{Y/X} = -\frac{dy}{dx} = \frac{y}{x}$, which is essentially equation (1). So, consider point A in Figure 3 with pool reserves (x^A, y^A). $P_{Y/X}$ offered by the CPMM can be read off the diagram in one of two equivalent ways: either as the absolute value of the slope of the tangent aa at A $(-\left.\frac{dy}{dx}\right|^A)$, or as the slope of the ray OA $(\frac{y^A}{x^A})$. Our view here is that an AMM is a black-box for converting inputs (token quantities) into outputs (prices). The equation $P_{Y/X} = \frac{y}{x}$ links the inputs in our black-box (x and y) to the output ($P_{Y/X}$) using the exchange function $k = xy$. Graphically, the output of the AMM is the slope of a level curve of the exchange function.

There are a couple of other issues about price determination in a CPMM that are worth noting. First, the exchange function $k = xy$ is homogenous and the slopes of level curves are the same along a ray from the origin. Since $P_{Y/X}$ is the absolute value of the slope, it follows that all points (reserve pairs) on a specific ray from the origin will entail the same price. Thus, returning to Figure 3, $P_{Y/X}$ is exactly the same at points A (where $k = k^0$) and C (where $k = k^1$), and indeed at every point along the ray OR, each of which falls on a different level curve. Secondly, when the amount of a token, say x, approaches zero, the slope of the level curve tends to infinity, and the token's price tends to infinity. Consequently, the reserves of a token

[32] The interested reader is referred to Silberberg (1990) or Chiang (1984) for a proof.

can never fall to zero due to trading activity as the level curves never intersect the axes.[33]

Next, we can see the price slippage that occurs with larger size transactions in Figure 3. At point C, the slope of the tangent cc determines $P_{Y/X}$ when an infinitesimal amount X is traded. For a larger trade that moves the CPMM from C to D, the price is given by the absolute value of slope of the line CD, which is larger than that of tangent cc. Similarly, when X is sold by the trader, the absolute value slope of the slope of CE is less than that of cc. It is readily visualized that larger the transaction size, greater is the slippage that occurs.

Figure 3 also describes what happens if liquidity providers were to add tokens to the pool. Starting from reserves (x^A, y^A), suppose liquidity providers add reserves in the same proportion as the initial reserves, so that $\frac{\Delta y}{\Delta x} = \frac{y^A}{x^A}$. Adding liquidity in the same proportion implies that new reserves (x^C, y^C) satisfy $\frac{y^C}{x^C} = \frac{y^A}{x^A}$; in other words, we move along ray OR, and prices are the same at the two points. Liquidity provision conducted in this manner changes the level curve from k^0 to k^1 without changing the price.

Consider now the process of two-point arbitrage, which is described in Figure 4 below.

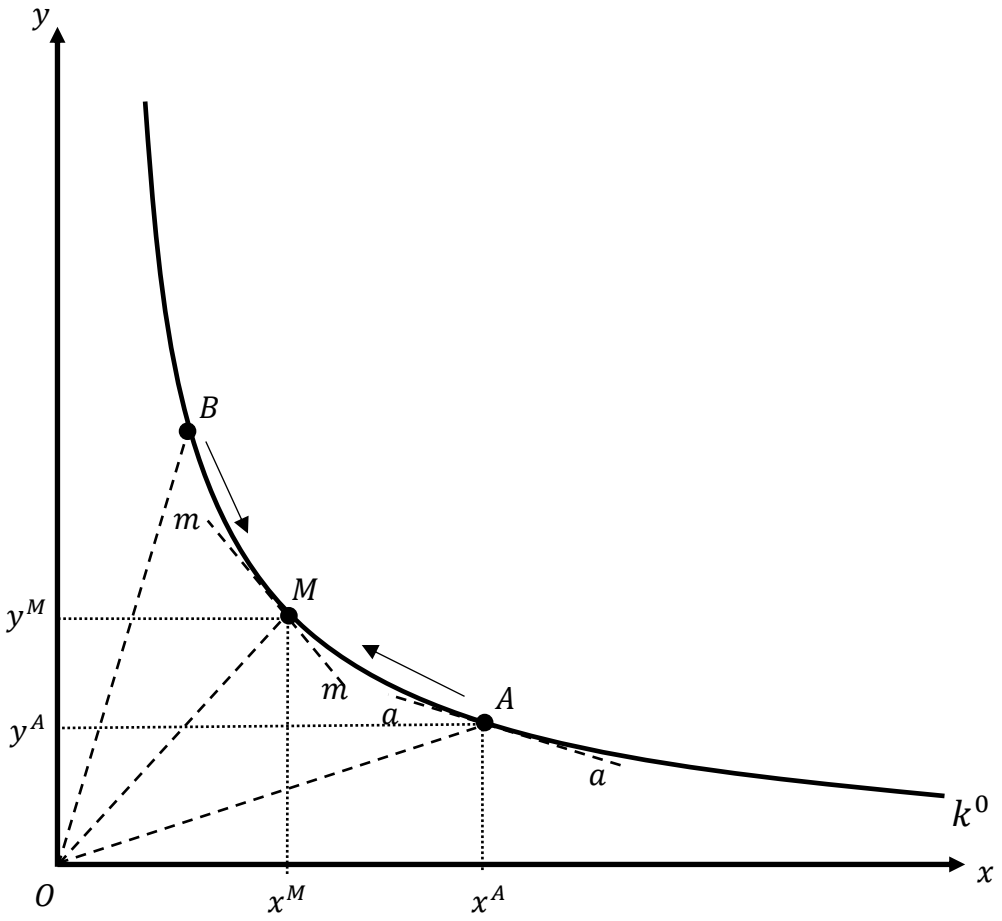

Figure 4: Price and arbitrage in a CPMM with no transaction fees

[33] Reserves can, of course, fall to zero if liquidity providers decide to withdraw their tokens.

Suppose the price of X in the external reference market is $M_{Y/X}$, where $M_{Y/X} > P_{Y/X}$. This external market price is shown in Figure 4 by the point M where the absolute value of the slope of the tangent mm equals the external market price. If the reserve levels in the CPMM is at point A, we have that $M_{Y/X} = -\frac{dy}{dx}\Big|^{M} > -\frac{dy}{dx}\Big|^{A} = P_{Y/X}$. This price differential triggers two-point arbitrage with arbitrageurs buying X from the CPMM and selling it in the external market. The buying activity in the CPMM increases $P_{Y/X}$ (while the selling in the external market can decrease $M_{Y/X}$), and the price in the CPMM will approach M, as shown by the arrow in Figure 4. This process continues till the no-arbitrage condition in equation (4) is met and $M_{Y/X} = P_{Y/X}$, which can occur anywhere between points A and M. When the external market is relatively large, the sale of X in the external market may have little impact on $M_{Y/X}$; in that case, $P_{Y/X}$ will settle at point M, with reserves (x^M, y^M). It is straightforward to construct the reverse argument if $M_{Y/X} < P_{Y/X}$ and the starting point is point B.

Now suppose there is a trading fee, τ. To examine this, we summarize the analytical procedure described in Angeris et al (2019) for deriving the bounds of arbitrage. Suppose we start with reserves (x^0, y^0), and $k^0 = x^0 y^0$. In the absence of trading fees, the price would be $P_{Y/X} = \frac{y^0}{x^0}$, irrespective of whether a trader wished to buy or sell X. The two-step process followed by Uniswap for price changes with a fee is described in detail in Section 2.1; we replicate that process analytically, assuming that it holds for any arbitrary CPMM.[34]

To begin with, assume that a trader wishes to sell X to the CPMM in exchange for Y. If the trader sells an amount Δx, the AMM retains $\tau\Delta x$, and the remainder $(1-\tau)\Delta x = \phi\Delta x$ is available for the trade, which occurs utilizing the initial value of the constant, k^0. If the trader receives Δy in return, we get:

(11) $$(x^0 + \phi\Delta x)(y^0 + \Delta y) = k^0$$

Since the liquidity pool receives X, we have that $\Delta x > 0$, from which it follows that $\Delta y < 0$. Simplifying, we get that $\frac{\Delta y}{\Delta x} = -\frac{\phi y^0}{x^0 + \phi\Delta x}$. When changes are infinitesimal, this reduces to $\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = -\phi \frac{y^0}{x^0}$. As $p^b_{Y/X} = -\frac{dy}{dx}$ in this case, we get the bid-price as:

(12) $$p^b_{Y/X} = \phi \frac{y^0}{x^0}$$

Similar arguments when the trader buys X from the CPMM ($\Delta y > 0$) yields an ask-price of:

(13) $$p^a_{Y/X} = \frac{1}{\phi}\frac{y^0}{x^0}$$

Given that that $\phi < 1$, it is evident that $p^b_{Y/X} < P_{Y/X} < p^a_{Y/X}$, where $P_{Y/X} = \frac{y^0}{x^0}$.

[34] Other CPMMs may adopt different procedures. However, the broad ideas that follow from homotheticity and so on described in this paper will not change due to small procedural changes.

Figure 5 below shows the bid and offer rate graphically. The reserves are depicted by point P, and the slope of ray OP is $\frac{y^0}{x^0} = P_{X/Y}$. This point lies on the level curve k^0. The slope of the ray OB is $\phi\frac{y^0}{x^0} = p^b_{Y/X}$; we can visually associate point B, therefore, with the bid-price. Similarly, the slope of OA is $\frac{1}{\phi}\frac{y^0}{x^0} = p^a_{Y/X}$, and point A is associated with the ask-price.

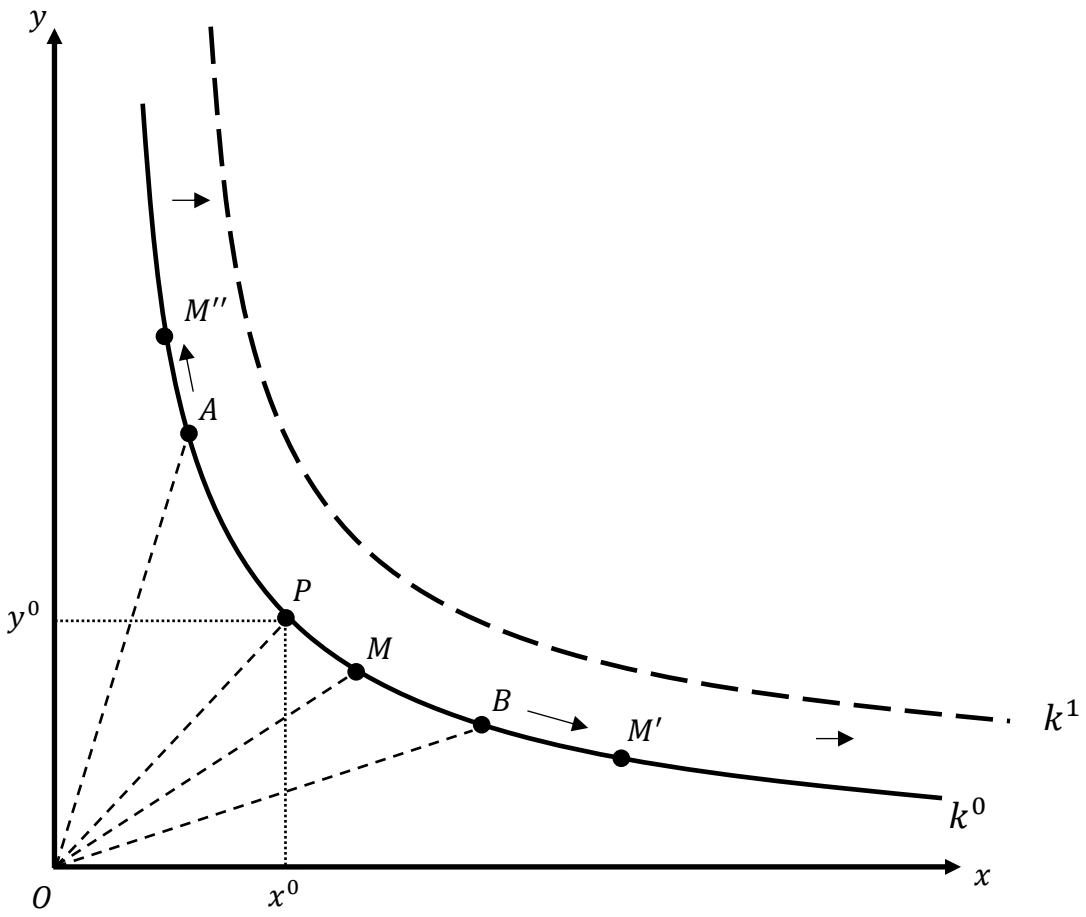

Figure 5: Price and arbitrage in a CPMM with transaction fees

To bring in the idea of arbitrage in Figure 5, suppose the external market price is $M_{Y/X}$. This is represented by a point M, which (uniquely on curve k^0) satisfies the condition that the slope of $OM = M_{Y/X}$. In this particular instance, M lies within the arc AB. As the slope of OM is greater than that of OB, but less than that of OA, we have that $p^b_{Y/X} \leq M_{Y/X} \leq p^a_{Y/X}$ holds at point M. This is, of course, nothing but the no-arbitrage condition in equation (5). This inequality is true for any point we select in the arc AB; thus, this arc corresponds to equation (5) as the range where arbitrage is not feasible. It is evident that the range widens (and the arc AB becomes longer) the greater the value τ takes, since $p^b_{Y/X} = \phi\frac{y^0}{x^0}$ and $p^a_{Y/X} = \frac{1}{\phi}\frac{y^0}{x^0}$. This goes back to the point made earlier: larger transaction costs curtail arbitrage.

If we consider a point outside arc AB, such as point M', where $p^b_{Y/X} > M_{Y/X}$, then

arbitrageurs can buy X in the external market and sell it in the CPMM to make profits. This sale of X to the CPMM will drive point P to the right, and the arc AB along with it, till $p^b_{Y/X} = M_{Y/X}$. Similar arguments in the reverse direction hold if $M_{Y/X}$ is represented by M''.

The second step of the exchange process in Uniswap-v1 and v2 is that the fees are added to generate a new value of k that satisfies $k^1 = (x^0 + \Delta x)(y^0 + \Delta y) = x^1 y^1$. In the case where X is sold to the CPMM, for example, $\tau \Delta x$ is added to the pool reserves, which causes a rightward shift of the level curve to k^1. The starting point for the next trade is then a reserve pair (x^1, y^1) on k^1, which has a no-trading fee price of $P_{X/Y} = \frac{y^1}{x^1}$ and a corresponding bid-ask spread as derived earlier.

3.3 Valuing a liquidity pool and a CPMM: Uniswap-v1 and v2

Consider a specific liquidity pool with tokens X and Y in CPMM-v1 or v2. There are circumstances where one may wish to value the liquidity pool. Since x and y are different tokens, they have to be brought to the same unit of account before their values can be added. Let us suppose that X (ETH) is the unit of account in the $X - Y$ pool. One way to value the liquidity pool is then $V_{X,Y} = x + P_{X/Y} y$.[35] This converts Y to an equivalent amount of X using $P_{X/Y} = \frac{x}{y}$. Substituting $P_{X/Y} = \frac{x}{y}$, it follows that:

(14) $\quad V_{X,Y} = 2x$

Equation (14) states that in order to value a liquidity pool that has ETH in it, one simply needs to double the amount of ETH in the pool. An implication of equation (14) is that the valuation of a pool is independent of the price $P_{X/Y}$ at which X and Y trade and, indeed, independent of the value of k. This is shown in Figure 6 below. All three points, A, B and C have a valuation of $2x^0$, even though they lie on different level curves and are associated with different prices. On the other hand, even though C, D and E have the same prices (they are on the same ray from the origin), the liquidity pool has different valuations at these points, with the valuation at C being the highest as it has the highest amount of X.

[35] See also Angeris et al (2019) and Angeris and Chitra (2020), who address the issue of pool valuation.

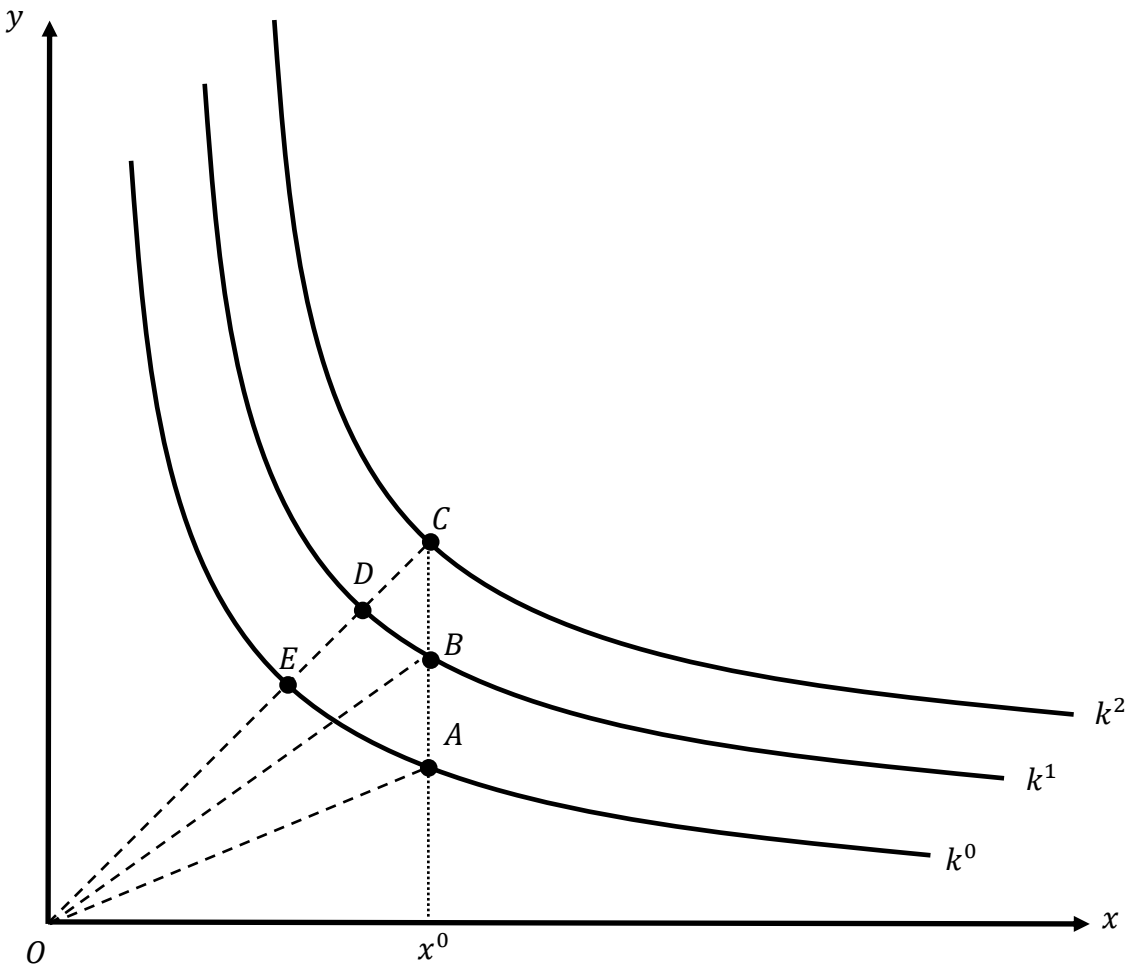

Figure 6: Liquidity pool valuation

This is, of course, not a unique way of valuing the pool. With trading fees and a bid-ask spread one may wish to use the mid-rate $\tilde{p}_{X/Y} = \frac{p^a_{X/Y}+p^b_{X/Y}}{2} = \frac{(1+\phi^2)}{2\phi}\frac{x}{y}$ (from equations (12), (13) and (3)) to value the pool. However, the mid-rate is a function of ϕ, which implies that the pool valuation using $\tilde{p}_{X/Y}$ to convert units of Y to X would be sensitive to the fees. Whether this is desirable or not depends on the context. The pool valuation in this case is:

(15) $\quad \tilde{V}_{X,Y} = x + \tilde{p}_{X/Y}y = \frac{(1+\phi)^2}{2\phi}x$

In the extreme case where there are no trading fees, $\phi = 1$ and it is evident that $\tilde{V}_{X,Y} = V_{X,Y}$. Moreover, as $\frac{d}{d\phi}\left(\frac{(1+\phi)^2}{2\phi}\right) < 0$, $\tilde{V}_{X,Y}$ falls as ϕ rises. This implies that as fees fall, the pool valuation decreases as well. This makes intuitive sense if one were to value the pool from the perspective of liquidity providers. Lower fees signal lower returns to pool providers, and $\tilde{V}_{X,Y}$ is lower as well. Thus, whether one wishes to use (14) or (15) depends on the perspective one takes. The valuation $V_{X,Y}$ is preferable if one looks at the liquidity pool as a whole. On the

other hand, if the focus is on liquidity providers, then $\tilde{V}_{X,Y}$ is a candidate for pool valuation.

One can generalize equation (14) to a broader class of functions where $k = f(x,y) = x^{\alpha}y^{\beta}$, with $\alpha, \beta > 0$. The $k = xy$ function is simply a special case of this where $\alpha = \beta = 1$. Using Euler's theorem from Section 3.1 for this case, we have that:

(16) $$x + \left(\frac{\partial f/\partial y}{\partial f/\partial x}\right) y = \frac{rf(x,y)}{\partial f/\partial x}$$

Since $\left(\frac{\partial f/\partial y}{\partial f/\partial x}\right) = -\frac{dx}{dy}$, it follows that $P_{X/Y} = \left(\frac{\partial f/\partial y}{\partial f/\partial x}\right)$.[36] The left-hand side of equation (16) is, therefore, nothing but $V_{X,Y}$. For the right-hand side, using $f(x,y) = x^{\alpha}y^{\beta}$, $\frac{\partial f}{\partial x} = \alpha x^{\alpha-1}y^{\beta}$, and $r = (\alpha + \beta)$, equation (16) reduces to:

(17) $$V_{X,Y} = \left(\frac{\alpha+\beta}{\alpha}\right) x$$

Comparing (17) with (14), we see that the former collapses to the latter whenever $\alpha = \beta$. So, doubling the amount of X (ETH) in a liquidity pool to value a liquidity pool is a valid technique for all AMMs that attach equal weights to X and Y in the $k = x^{\alpha}y^{\beta}$ function, and is not limited to the specific exchange function used by Uniswap. For this general class of exchange functions, even when $\alpha \neq \beta$, the pool valuation depends only on x and some constant factor $\left(\frac{\alpha+\beta}{\alpha}\right)$.[37] It is also worth noting that the share of value of X to the value of the pool is $x/V_{X,Y}$, which equals $\frac{\alpha}{\alpha+\beta}$ and is fixed once the parameters α and β are fixed.[38] Similarly, the share of Y is $P_{X/Y}y/V_{X,Y} = \frac{\beta}{\alpha+\beta}$. When $\alpha = \beta$, each token's share of the total pool value is ½.

Now suppose we would like to value an entire CPMM operating with an exchange function $k = xy$ across all liquidity pools that exist on the platform. In a CPMM-v1, as outlined earlier, all liquidity pools have ETH as a common token. If there are n liquidity pools, each of which has X as token, along with another token, $Y_i, i \in \{1,2,3,\dots n\}$. The n liquidity pools yield a set of reserves: $\{(x_1,y_1),(x_1,y_1)\dots(x_n,y_n)\}$, where x_i is the amount of X (ETH) in the liquidity pool with token Y_i. From (14), each liquidity pool has a valuation $V_{X,Y_i} = 2x_i, \forall i \in \{1,2,3,\dots n\}$. This allows a simple method to value the entire CPMM-v1:

(18) $$V_{CPMM-v1} = \sum_{i=1}^{n} V_{X,Y_i} = 2\sum_{i=1}^{n} x_i$$

Equation (18) states that to value a CPMM-v1 with the exchange function $k = xy$, one needs to simply add all the ETH cross the various liquidity pools on the platform and double it.

Valuing CPMM-v2 is slightly more complicated, because there are liquidity pools available

[36] It is evident that $P_{X/Y} = \frac{\beta}{\alpha}\frac{x}{y}$ in this case.

[37] For the alternative $\tilde{V}_{X,Y}$, we have $\tilde{V}_{X,Y} = x + \tilde{p}_{X/Y}y = x + \frac{(1+\phi^2)}{2\phi}\frac{\beta}{\alpha}x = (1 + \frac{(1+\phi^2)}{2\phi}\frac{\beta}{\alpha})x$.

[38] Later on, in Section 4.1, we show that other types of AMMs, such as a Constant Mean Market Maker, also have a similar property of fixed shares of value.

that do not involve ETH. So apart from the n original liquidity pools with reserves $\{(x_1, y_1), (x_1, y_1) \dots (x_n, y_n)\}$ that were available in CPMM-v1, CPMM-v2 allows for an additional $\frac{n(n-1)}{2}$ liquidity pools between tokens Y_i and Y_j, $\forall i, j \in \{1,2, \dots n\}, i \neq j$. In terms of notation, let y_{ij} denote the quantity of Y_i in a $Y_i - Y_j$ liquidity pool. To value a liquidity pool with Y_i and Y_j, we would first to pick a pricing token (say Y_i); it follows from a parallel to equation (14) that $V_{Y_i,Y_j} = 2y_{ij}$. However, this is denominated in units of token Y_i; to convert this to units of X (ETH), so that there exists a common unit of account to value different pools, we can use the price $P_{X/Y_i} = \frac{x_i}{y_i}$ generated by the reserves (x_i, y_i) in the $X - Y_i$ liquidity pool. Thus, we get $V'_{Y_i,Y_j} = V_{Y_i,Y_j} \times P_{X/Y_i} = 2\frac{x_i}{y_i} y_{ij}$, which is denominated in X.

Suppose we use an indicator $d_{ij} = 1$ if Y_i is used as the pricing token in the pool with Y_j and $d_{ij} = 0$ if it is not. The value of all pools $Y_i - Y_j$ that uses Y_i as the pricing currency, but denominated in units of X is $V'_{Y_i} = 2P_{X/Y_i} \sum_{j \neq i} d_{ij} y_{ij}$. The total value of the CPMM-v2 is then:

(19) $V_{CPMM-v2} = \sum_{i=1}^{n} \left[2P_{X/Y_i} \sum_{j \neq i} d_{ij} y_{ij}\right] + 2\sum_{i=1}^{n} x_i$, where $P_{X/Y_i} = \frac{x_i}{y_i}$

The second term in (19) corresponds to equation (18), and refers to the value of all pools that involve X. The first term adds the values of all $Y_i - Y_j$ pools after denominating them in X.

3.4 Measuring liquidity in a CPMM

3.4.1 Interpreting k

How exactly is one to interpret the economic meaning of k in the exchange function $k = f(x_1, x_2, \dots, x_n)$? In consumer theory, k is interpreted as the level of utility, while in production theory, it is the quantity of goods and services produced. In an AMM, the 'output' is the price of one token in terms of the other, which is slope of the level curves of $f(.)$; this corresponds to the marginal rate of substitution in consumer theory and to the marginal rate of technical substitution in production theory. There is no direct economic significance of k that is apparent when calculating these slopes in the context of AMMs.[39]

Geometrically, we can associate a value of k with a level curve generated by the exchange function, as we have done in Figure 3. This affords the most satisfactory interpretation that one can ascribe to k - as an ordinal measure of *liquidity*, in much the same way as utility serves as an ordinal measure of satisfaction in consumer theory. Viewed in this way, we can think of the level curves of the exchange function as 'isoliquidity' curves.

Thus, if $k^1 > k^0$ in a given liquidity pool, it is intuitive to infer that the liquidity level at k^1 is higher than the liquidity level at k^0 because, starting at any point on k^0, the injection of

[39] Alternatively, in the case where, $k = xy$, as in Figure 3, we can think of k geometrically as the area of the rectangle associated with a point; thus at point A in Figure 3, we have that Area$(Ox^A Ay^A) = k^0$. However, the dimension of the area is units of X × units of Y, which has little economic meaning as well.

tokens by liquidity providers will move the reserves to a higher level curve like k^1. This provides a valid way of comparing the liquidity levels provided by any two arbitrary reserve token pairs (x^0, y^0) and (x^1, y^1) in a liquidity pool even if, say, $x^0 < x^1$ but $y^0 > y^1$; (x^1, y^1) can be said to provide a higher level of liquidity if $x^1 y^1 = k^1 > k^0 = x^0 y^0$.

The nature of the concept of liquidity as defined here suggests that inter-pool comparisons of k are meaningless. So, if we have two liquidity pools with different tokens and calculate that $x^1 y^1 = k^1 > k^0 = x^0 z^0$, we cannot conclude that the $X - Z$ pool has a lower level of liquidity than the $X - Y$ pool. Essentially, the numbers k^1 and k^0 are not comparable in this instance because their units of measurement are different.

3.4.2 Liquidity in a CPMM

While the parameter k is a measure of liquidity in a pool, it has a few problems, arising from the fact the function $k = xy$ is homogeneous of degree two. To see the implications of this, suppose one liquidity provider doubles the amount of tokens in the pool. The new value of the invariant is $4xy = 4k$; in other words, the invariant has quadrupled even though 'liquidity' has, for all practical purposes, only doubled due to the doubling of token volume. This is particularly problematic in CPMM-v1 and CPMM-v2, where fungible ERC-20 liquidity tokens are issued. If these tokens are issued one-to-one with the value of k, by doubling the tokens the liquidity provider in this example would have received 75% of the liquidity tokens in circulation, even though the provider has accounted for only half the tokens.

To quantify liquidity in general, and to keep track of liquidity tokens in particular, a better measure would involve using a function that is homogenous of degree 1. To that end, suppose we define $L = \sqrt{k} = x^{\frac{1}{2}} y^{\frac{1}{2}}$. Now, if a liquidity provider, for example, doubles both tokens, the value of L also doubles. Indeed, we can generalize the benefits of utilizing L to measure liquidity even for small additions to the liquidity pool. As shown in Section 2.3, in order for liquidity provision not to change price in CPMM-v1 and v2, the condition $\frac{dy}{y} = \frac{dx}{x}$ holds. It is readily seen that $\frac{dL}{L} = \frac{dx}{x}$ also holds,[40] implying that the percentage change in L equals the percentage change in either token. If X is ETH, then CPMM-v1 and v2 can distribute new liquidity tokens based on changes in L; specifically, $dL = \frac{dx}{x} \times L$ (or its discrete counterpart, $\Delta L = \frac{\Delta x}{x} \times L$). Consequently, L serves as a better measure of liquidity in CPMMs.

3.5 Competing interests in an AMM and the efficiency-liquidity provision trade-off

Liquidity providers act as the counterparty to all trading activity on an AMM. In this section we explore a number of issues related to liquidity provision and the conflicts of interests that

[40] Taking log of $L = x^{\frac{1}{2}} y^{\frac{1}{2}}$ and differentiating yields $\frac{dL}{L} = \frac{1}{2}\frac{dx}{x} + \frac{1}{2}\frac{dy}{y}$. Since $\frac{dy}{y} = \frac{dx}{x}$, it follows that $\frac{dL}{L} = \frac{dx}{x}$.

can arise in this context. In order to do so, we keep track of four types of agents who participate in an AMM: liquidity providers, traders interested in swapping assets, arbitrageurs, and attackers who exploit the system. While we approach this in the context of a CPMM, the general principles hold for other types of AMMs, even if specifics may vary.

Consider traders to begin with. Liquidity providers earn returns as fees from transactions with traders, even if there were no movements in the external market price. Clearly, traders would prefer the trading fee to be as little as possible. From the point of view of a liquidity provider, a higher fee yields higher returns per trade, but discourages trading activity. To characterize this trade-off in a simple manner, suppose $v(\tau)$ is the (average) value of tokens transacted by traders on the AMM (denominated in a numeraire, say X) over a given period of time, where τ is the *ad valorem* fee. Since higher fees reduce trading activity, it is reasonable to suppose that $\frac{dv}{d\tau} < 0$. Assuming liquidity provision costs are normalized to zero, the profits that liquidity providers get from traders is $\tau v(\tau)$. To maximize this, liquidity providers would set the optimum fee, τ^*, which satisfies the first order condition $\frac{\tau^* \frac{dv(\tau^*)}{d\tau}}{v(\tau^*)} = -1$. In terms of interaction with traders, the AMM essentially provides a service, and the optimum fees from the point of view of liquidity providers balances the trade-offs between the price of the service (τ) and the demand for the service ($v(\tau)$).

However, liquidity providers also have to contend with arbitrageurs. In Figure 5, traders will participate even when the external market price lies within the arc AB, but arbitrage is not profitable in that range. When the external market price moves beyond AB, arbitrageurs are vital for AMMs such as Uniswap to achieve an alignment between the AMM price and the external market price. If trading fees are increased to favor liquidity providers, the arc AB widens, which discourages arbitrage over a wider range of prices. This indicates an intuitive trade-off between efficiency in terms of price alignment with external markets, and the incentives for liquidity provision.

It is worthwhile exploring further how external market price movements and the resultant arbitrage activities affect liquidity providers. To do so, we assume that there are no traders, so that there is only interaction between liquidity providers and arbitrageurs. To begin with, consider the case where $\tau = 0$; Figure 4 serves as a visual reference. One way to measure the performance of liquidity provision is to compare it to a situation when the same assets are held in the external market. Essentially, this is the opportunity cost of placing the tokens in the CPMM, assuming that the next best alternative is holding it in an external wallet.

At the initial equilibrium when $P^0_{Y/X} = M^0_{Y/X}$ (say point A in Figure 4), the reserves in a CPMM are (x^0, y^0). Using the external price to value assets, the value of tokens in the CPMM liquidity pool (measured in units of Y) is $v^0_C = y^0 + M^0_{Y/X} x^0$, which equals the value in the external market, v^0_E. Now suppose the price in the external market were to change to $M^1_{Y/X} >$

$M^0_{Y/X}$ (say point M in Figure 4). Arbitrage implies that we move from point A to M, so that $P^1_{Y/X} = M^1_{Y/X}$ and new CPMM reserves are (x^1, y^1). The values of the tokens in the two cases are now $v^1_C = y^1 + M^1_{Y/X}x^1$ and $v^1_E = y^0 + M^1_{Y/X}x^0$. Even though $P^1_{Y/X} = M^1_{Y/X}$, the values are different because the CPMM has *rebalanced* to ensure that $y^1 = P^1_{Y/X}x^1$, whereas no such automatic rebalancing occurs in the external market. Since the arbitrage activity implies that CPMM liquidity providers are holding less of the token that is gaining value (X in this case), it is intuitive that $v^1_C < v^1_E$ will hold. Letting $\theta = \frac{M^1_{Y/X}}{M^0_{Y/X}} = \frac{P^1_{Y/X}}{P^0_{Y/X}}$, the percentage difference in the CPMM pool value relative to the external market value when there are no fees is:[41]

(20) $$I(\theta) = \frac{v^1_C - v^1_E}{v^1_E} = \frac{2\sqrt{\theta}}{1+\theta} - 1$$

Since $2\sqrt{\theta} < (1+\theta)$, we have that $I(\theta) < 0$, and liquidity providers suffer a loss in value relative to holding a non-rebalanced portfolio in the external market. Moreover, as this inequality holds for any price ratio θ, the movement in the reverse direction (where say the change is from point B to M in Figure 4) also yields similar results. In the DeFi community, $I(\theta)$ is referred to as *impermanent loss* or *divergence loss*, because this is an unrealized loss relative to the external market; if the price were to revert to $M^0_{Y/X}$ again, we would have that $v^0_C = v^0_E$, and the loss disappears. As this primarily reflects the impact of rebalancing in the CPMM, one could relabel this as a *rebalancing cost*.

Adding fees moves the analysis from Figure 4 to Figure 5, which introduces additional features: first, there exists a range (the arc AB) where no arbitrage takes place; and second, there are multiple prices in the CPMM ($P_{Y/X}, p^b_{Y/X} = \phi P_{Y/X}$ and $p^a_{Y/X} = \frac{1}{\phi}P_{Y/X}$). To see how these features play out, suppose the initial equilibrium is at point P in Figure 5, so that $M^0_{Y/X} = P_{Y/X}$. Now, if the external price reduces to $M^1_{Y/X}$ equal to slope of ray OM, no arbitrage is triggered and the CPMM reserve assets continue to be located at P. In this case no rebalancing occurs, and the valuation exercise is trivial: if the same price is used for valuing assets in both cases, there is no difference in the valuation of tokens.

Things become more meaningful once we reach the boundaries of the arc AB. Suppose the initial equilibrium is at B, where $M^0_{Y/X} = p^{b0}_{Y/X}$. Now, a small fall in the external price to $M^1_{Y/X} < M^0_{Y/X}$ (point M') will trigger arbitrage till the new bid price satisfies $p^{b1}_{Y/X} = M^1_{Y/X}$.[42]

[41] See Pintail (2019) for this formula, along with examples. Formally, from the relationships $xy = k$ and $\frac{y}{x} = P_{Y/X}$, we have that $x = \sqrt{\frac{k}{P_{Y/X}}}$ and $y = \sqrt{kP_{Y/X}}$. This holds both before and after the price change, along with the fact that $M_{Y/X} = P_{Y/X}$ at equilibrium. So, $v^1_C = 2\sqrt{\theta k P^0_{Y/X}}$ and $v^1_E = (1+\theta)\sqrt{kP^0_{Y/X}}$; it follows that $I(\theta) = \frac{2\sqrt{\theta}}{1+\theta} - 1$. See Xu et al (2021) for calculations for other types of AMMs.

[42] We assume that the external market is large enough that arbitrage with the CPMM does not affect $M_{Y/X}$.

On the other hand, a small rise in price will not result in arbitrage, as we return to the interior of arc AB.[43] Consequently, the useful case to examine is $\theta = \frac{M^1_{Y/X}}{M^0_{Y/X}} = \frac{p^{b1}_{Y/X}}{p^{b0}_{Y/X}} < 1$. The valuations are now $v^1_C = y^1 + M^1_{Y/X}x^1$ and $v^1_E = y^0 + M^1_{Y/X}x^0$, where $M^1_{Y/X} = p^{b1}_{Y/X}$, and so:[44]

(21) $$I(\theta, \phi) = \frac{v^1_C - v^1_E}{v^1_E} = \frac{\sqrt{\theta}(1+\phi)}{1+\theta\phi} - 1$$

It follows that $I(\theta, \phi) < 0$ when $\theta < \frac{1}{\phi^2}$, which must hold when $\theta < 1$. As a result, fees do not alter the fact that rebalancing costs exist beyond the range of the spread. A similar analysis occurs when $M^0_{Y/X} = \ p^{a0}_{Y/X}$ and $\theta > 1$; in this case $I(\theta, \phi) = \frac{v^1_C - v^1_E}{v^1_E} = \frac{\sqrt{\theta}(1+\phi)}{\phi+\theta} - 1$.[45]

Our comparative static analysis suggests that the costs of rebalancing are inevitable, and liquidity providers must depend on fees from users to outweigh this. However, this does not factor in how price movements impact the compounding of wealth over time. Tassy and White (2020) show that, under certain circumstances, rebalancing is useful to minimize the negative effect of losses on compounding. In this scenario, it can be optimal to set the fees as low as possible (without being zero), thereby taking advantage of rebalancing when prices are volatile around a narrow spread.[46] While this presents some preliminary insights into the dynamics in a specialized circumstance, there is more work to be done in this area.

Fees can impact the security of the system in terms exploits, such as a sandwich attack.[47] A sandwich attack involves both front-running and back-running a transaction. Suppose a trader places an order to sell Y and buy X. Upon seeing the (public) order, an attacker can front-run the transaction by buying X at the offer price ($p^{a0}_{Y/X}$), thereby driving up the price of X. The trader's original transaction drives the price even higher,[48] and the attack is then completed with a back-run sale of X by the attacker at a new bid price $p^{b1}_{Y/X}$; the profit of the attacker is $p^{b1}_{Y/X} - p^{a0}_{Y/X}$. For (no-fee) prices $P^1_{Y/X}$ and $P^0_{Y/X}$, the profit equals $\phi P^1_{Y/X} - \frac{1}{\phi} P^0_{Y/X}$, which is decreasing in the trading fee τ. A higher trading fee can, therefore, reduce

[43] A large increase in price, of course, can trigger arbitrage if the new price is higher than slope OA. But recalling that these price formulas are for small changes, we focus on this case.

[44] There are now three relationships we work with: $p^b_{Y/X} = \phi P_{Y/X}$, $P_{Y/X} = \frac{y}{x}$ and $xy = k$. These yield $x = \sqrt{\frac{\phi k}{p^b_{Y/X}}}$ and $y = \sqrt{\frac{k\, p^b_{Y/X}}{\phi}}$, which hold both before and after the price change. The valuations are then $v^1_C = (\frac{1+\phi}{\sqrt{\phi}})\sqrt{\theta k\, p^{b0}_{Y/X}}$ and $v^1_E = (\frac{1+\theta\phi}{\sqrt{\phi}})\sqrt{k\, p^{b0}_{Y/X}}$. It follows that $I(\theta) = \frac{v^1_C - v^1_E}{v^1_E} = \frac{\sqrt{\theta}(1+\phi)}{1+\theta\phi} - 1$.

[45] We have that $\frac{\sqrt{\theta}(1+\phi)}{\phi+\theta} < 1$ when $\theta > \phi^2$, which must be true for the relevant range where $\theta > 1$.

[46] See White et al (2020) for an intuitive description of these ideas.

[47] See Zhou et al (2020) for a deeper analysis of sandwich attacks. See also Buterin (2018) for an early exploration.

[48] As a practical matter, the exact amount by which price changes depends on the extent of the slippage; see Zhou et al (2020).

the incentives for such attacks and favor traders whose transactions are sandwiched (though the higher fees increase the spread, which affects traders adversely).

Intuition suggests that the nature of the tokens being traded is also important when determining fees. As such, tokens with low volatility and stable relative prices (for example, correlated token pairs) can afford a lower width of the arc AB (that is, lower fees) in Figure 5 compared to more volatile assets. The low fees encourage more trading activity in low-volatility tokens, while the price stability minimizes the need for rebalancing due to arbitrage activity, even with a low spread. On the other hand, for tokens that exhibit larger relative price movement, a narrow arc AB would result in frequent arbitrage and rebalancing. In that circumstance, it may be worthwhile to impose higher fees.[49]

The design of AMM fees must attempt to balance all these interests. Higher trading fees benefit liquidity providers and reduce exploits, but discourage trading and arbitrage, thereby reducing efficiency. Lower fees have a reverse effect. Overall, the design of optimum fees for an AMM is complicated and would have to weigh all these trade-offs. Current research has, in general, focussed on the optimizing fees for a single entity (typically, liquidity providers, as in Tassy and White, 2020). As such, this is an ongoing research endeavour.

4. Other types of AMMs for decentralized exchanges

In this section, we introduce other types of AMMs. While these are not covered in nearly the same detail as the CPMM, many of the techniques and insights from Sections 2 and 3 carry over here, so that they can be extrapolated in a straightforward way.

4.1 Constant Mean Market Makers (CMMM)

A constant mean market maker (CMMM) has reserves $(x_1, x_2, \dots x_n)$ of n tokens $X_1, X_2, \dots X_n$ that satisfy the exchange function:

(22) $\prod_{i=1}^{n} x_i^{w_i} = k$, where $\sum_{i=1}^{n} w_i = 1$

The CPMM is a special case of (22) where $n = 2$ and $w_i = 0.5$ for $i \in \{1,2\}$, which can then be squared to yield a CPMM of the type examined before. An example of a DeFi platform which operates as a CMMM is Balancer,[50] and our analysis here is related to the Balancer whitepaper (Martinelli and Mushegian, 2019). An important aspect of the Balancer protocol is that the weights, once fixed, determine the share of the value a particular token has in relation to the overall value of the pool; from this point of view, the Balancer protocol has similarities to index funds that construct a portfolio of assets with fixed weights to each asset

[49] Indeed, Uniswap-v3 allows for multiple fee structures, initially set at 0.05% (intended for stablecoins), 0.3% (for standard token pairs) and 1% (for highly volatile tokens).
[50] See https://balancer.finance (accessed on 20th September, 2021).

(Martinelli and Mushegian, 2019).[51]

To begin the analysis, suppose there are no trading fees. The bilateral trading price of one unit of X_i in terms of X_j is $P_{X_j/X_i} = -\frac{dx_j}{dx_i}$. Letting $\prod_{i=1}^{n} x_i^{w_i} = f(x_1, x_2, \dots x_n)$, equation (22) yields $\sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i = dk$. Along any particular level curve $dk = 0$; moreover, since only quantities x_i and x_j change in this transaction, $dx_l = 0$ for $l \neq \{i, j\}$. Consequently, for any arbitrary reserve quantities $(x_1, x_2, \dots x_n)$, $\frac{dx_j}{dx_i} = -\frac{\partial f/\partial x_i}{\partial f/\partial x_j} = -\frac{w_i}{w_j}\frac{x_j}{x_i}$, and we have:

(23) $$P_{X_j/X_i} = \frac{w_i}{w_j}\frac{x_j}{x_i}$$

We are now in a position to see how the process of two-point arbitrage works in a CMMM. Consider the case where $M_{X_j/X_i} < P_{X_j/X_i} = \frac{w_i}{w_j}\frac{x_j}{x_i}$. Arbitrageurs will purchase X_i in the external market and sell it in the CMMM in exchange for X_j. This causes $\frac{x_j}{x_i}$ to fall till the no-arbitrage condition $M_{X_j/X_i} = P_{X_j/X_i}$ is restored.

We can also calculate the liquidity pool value after selecting a numeraire to do the pricing. Without loss of generality, let us suppose X_1 is selected as the token to price the liquidity pool; X_1 could be, for example, ETH. Using Euler's theorem:[52]

(24) $$V_{X_1, X_2, \dots, X_n} = \frac{x_1}{w_1}$$

The share of any token value to the pool value is:

(25) $$s_i = \frac{x_i P_{X_1/X_i}}{V_{X_1, X_2, \dots, X_n}} = w_i, \forall i \in \{1, 2, \dots n\}$$

In a CMMM, therefore, the share of the value of any token to the value of the entire pool is fixed and equal to the weight of each token in the liquidity pool, which is similar to a CPMM.

Now suppose we add *ad valorem* fees, τ, to the price calculation problem assuming, as before, that the fees are paid in units of the token sold by a trader into the liquidity pool. While slightly more tedious to calculate, we can show that $p^b_{X_j/X_i} = \phi \frac{w_i}{w_j}\frac{x_j}{x_i}$ and $p^a_{X_j/X_i} = \frac{1}{\phi}\frac{w_i}{w_j}\frac{x_j}{x_i}$.[53] Using equation (5), the no-arbitrage condition is:

---

[51] As shown earlier, a CPMM also shares this feature.

[52] From Euler's theorem we have that $x_1 + \frac{\partial f/\partial x_2}{\partial f/\partial x_1} x_2 + \cdots \frac{\partial f/\partial x_n}{\partial f/\partial x_1} x_n = \frac{rf(x_1, x_2, \dots x_n)}{\partial f/\partial x_1}$. Since $\frac{\partial f/\partial x_i}{\partial f/\partial x_1} = \frac{w_i}{w_1}\frac{x_1}{x_i} = P_{X_1/X_i}$, the left hand side of Euler's equation is $V_{X_1, X_2, \dots, X_n}$. Given the constraint that $\sum_{i=1}^{n} w_i = 1$, we have that $r = 1$, and the right-hand side equals $\frac{x_1}{w_1}$.

[53] Suppose the trader sells X_i and purchases X_j; this yields the bid-price $p^b_{X_j/X_i}$. Assume we start with pool reserves $(x_1, x_2, \dots, x_n)$ which satisfies $\prod_{i=1}^{n} x_i^{w_i} = k$. For the trade we have $\left(\prod_{\substack{r=1 \\ r \neq i,j}}^{n} x_r^{w_r}\right)(x_i + \phi \Delta x_i)^{w_j}\left(x_j + \Delta x_j\right)^{w_j} = \prod_{i=1}^{n} x_i^{w_i} = k$. This implies that $\Delta x_j = x_j\left[\left(1 + \frac{\phi \Delta x_i}{x_i}\right)^{-\frac{w_i}{w_j}} - 1\right]$, from which it follows that $\lim_{\Delta x_i \to 0} \frac{\Delta x_j}{\Delta x_i} =$

(26) $\phi \frac{w_i}{w_j}\frac{x_j}{x_i} \leq M_{X_j/X_i} \leq \frac{1}{\phi}\frac{w_i}{w_j}\frac{x_j}{x_i}$

4.2 Constant Sum Market Makers (CSMM)

There has been some discussion in the DeFi space about the usefulness and limitations of a constant sum market maker (CSMM).[54] For reserves (x, y), a CSMM holds the sum of reserves constant, that is, the exchange function satisfies $x + y = k$. Of course, one could generalize this to $ax + by = k$, or to $a_1x_1 + a_2x_2 + \cdots + a_nx_n = k$.

The main point we wish to make here is that a CSMM functions poorly as a decentralized exchange because arbitrage activity can entirely drain the liquidity pool of a token, and the simple $x + y = k$ case suffices to show that. Before doing so, however, it is worthwhile examining why a CSMM may be attractive. One of the features of a CPMM (and a CMMM) is that along a particular level curve, say k^0 in Figure 4, the slope varies at every point, and consequently the price changes every time the reserves change, even if by a small amount. In contrast, in a CSMM $ax + by = k$, the price is constant and equal to $P_{Y/X} = -\frac{dy}{dx} = \frac{a}{b}$; indeed, in the special case where $a = b = 1$, $P_{Y/X} = 1$, and the two tokens trade on a one-to-one basis. The price stability is attractive in that it reduces slippage, which is a useful feature to have when trading tokens with stable relative prices.

In a decentralized exchange, however, the constant price in a CSMM is problematic if the external market price is variable. Intuitively, arbitrageurs will constantly take advantage of the price differential in the two markets and because the constant price in the CSMM cannot, by construction, respond to this pressure by arbitrageurs, they may drain the CSMM of one token or the other. This intuition is described more systematically in Figure 7, which depicts the exchange function $x + y = k$. Suppose that $\tau = 0$ to begin with.

Consider a situation where the reserves in the CSMM are (x^A, y^A), and the level curve is $k = k^0$. The absolute value of slope of the line BD is $P_{Y/X} = 1$. Now suppose the external market price is $M_{Y/X} > 1$. In this case, arbitrageurs will buy X in the CSMM and sell it in the external market to make a profit of $M_{Y/X} - P_{Y/X}$ for every unit of X bought and sold this way. If the CSMM is small compared to the external market, $M_{Y/X}$ may change little in response to the sale of X there, and the process will continue till the CSMM is drained of all X since $P_{Y/X}$

$\lim_{\Delta x_i \to 0} \frac{x_j\left[\left(1+\frac{\phi \Delta x_i}{x_i}\right)^{-\frac{w_i}{w_j}}-1\right]}{\Delta x_i}$. Using L'Hôpital's rule, the right-hand side is $-\phi \frac{w_i}{w_j}\frac{x_j}{x_i}$, and consequently, $p^b_{X_j/X_i} = -\frac{dx_j}{dx_i} = \phi \frac{w_i}{w_j}\frac{x_j}{x_i}$. The proof for $p^a_{X_j/X_i} = \frac{1}{\phi}\frac{w_i}{w_j}\frac{x_j}{x_i}$ follows along similar lines.

[54] An example of using a CSMM as the exchange function is mStable, which used a CSMM to initially develop mUSD. However, as users identified concerns with the design, it has moved away from the CSMM model for mBTC: see https://github.com/mstable/MIPs/blob/master/MIPS/mip-7.md (accessed 20th September, 2021). Egorov (2020) discusses the problems with a CSMM, before utilizing a mix of the CSMM and CPMM exchange functions to develop a hybrid AMM.

is fixed.

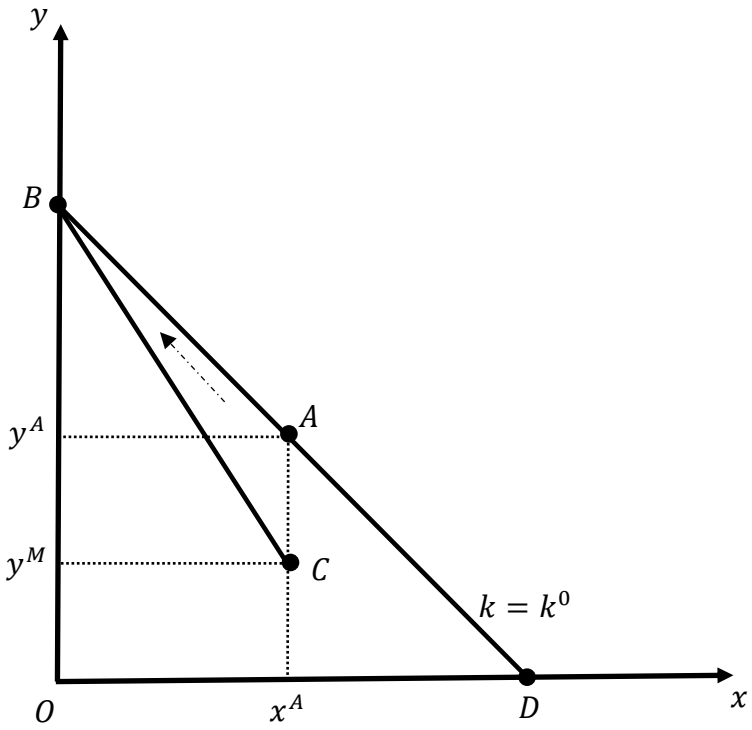

Figure 7: Price and arbitrage in a CSMM with no trading fees

In Figure 7, starting from point A, arbitrageurs can buy Ox^A units of X by selling By^A units of Y to the CSMM, which causes a movement to the corner point B. The line BC describes the trading possibilities in the external market, and has a slope greater than 1, thereby capturing the fact that $M_{Y/X} > P_{Y/X}$. Arbitrageurs can sell the Ox^A units of X purchased in the CSMM for By^M units of Y in the external market, making an overall profit of $y^M y^A$ units of Y. For the CSMM, the arbitrage process drains all X reserves and the CSMM is left with k^0 units of Y, leading to a corner equilibrium at point B.[55] The arguments when $M_{Y/X} < 1$ are similar, with a corner equilibrium at D.

Adding trading fees as before, we have that $p^b_{Y/X} = \phi$ and $p^a_{Y/X} = \frac{1}{\phi}$. In the scenario above where an arbitrageur buys X from the CPMM, the relevant price of X is $\frac{1}{\phi}$ units of Y. So, if fees are high and $\frac{1}{\phi} > M_{Y/X}$, arbitrage is no longer profitable; on the other hand, if fees are

[55] A corner solution such as this is discussed in many undergraduate microeconomics textbooks (see, for example, Varian, 1987) in the context of perfect substitution between goods in a utility function. In this case, indifference curves are straight lines. The consumer simply spends her entire budget on the lower priced good, yielding a corner solution to the consumer utility maximization problem. The fact that the AMM will lose the more valuable token in its reserves also bears some similarity in a macroeconomic context to Gresham's Law: bad money drives out good money (for example, see Dernburg, 1989). This phenomenon typically happens when there are multiple competing reserves, for example, gold and silver in a bimetallic standard.

low and $\frac{1}{\phi} < M_{Y/X}$, arbitrage continues as before and yields a corner equilibrium. Overall, the no-arbitrage condition in this case is $\phi \leq M_{Y/X} \leq \frac{1}{\phi}$. Higher the fees, wider is the gap and it is less likely that the CSMM is depleted of one token due to the action of arbitrageurs.

4.3 Hybrid Function Market Makers (HFMM)

One of the advantages of a CSMM is that there is no price slippage, which is a problem with a CPMM/CMMM. However, the disadvantage of a CSMM is the possibility of a corner equilibrium when token prices are volatile, which cannot happen in a CPMM/CMMM. A mix of a CPMM/CMMM and a CSMM that emphasizes both their advantages, while minimizing their problems, would be a desirable middle-ground in many situations, such as the exchange of stablecoins. A Hybrid Function Market Maker (HFMM) attempts to achieve this by integrating the exchange functions of a CSMM and a CPMM/CMMM.

Much of the discussion below is inspired by the whitepaper for Curve Finance (or, formerly, Stableswap; see Egorov, 2019), an AMM that facilitates trade in stablecoins. Our focus here is to bring out the intuition behind how an integration of a CPMM/CMMM and CSMM may be achieved and the HFMM described here is, consequently, considerably less complex than that of Egorov (2019).[56] To keep the exposition simple, trading fees are assumed to be zero.

Consider a CMMM described in equation (22), with equal weights so that $\prod_{i=1}^{n} x_i^{\frac{1}{n}} = k$ and a CSMM $a_1x_1 + a_2x_2 + \cdots + a_nx_n = k$, also with equal coefficients $a_1 = a_2 = \cdots = a_n = \frac{1}{n}$, which yields $\frac{\sum_{i=1}^{n} x_i}{n} = k$. Why these special cases with equalized weights and coefficients were chosen will become apparent presently. The bilateral price in the CMMM is $P_{X_j/X_i} = \frac{x_j}{x_i}$ and in the CSMM is $P_{X_j/X_i} = 1$. Our goal is to mix these two AMMs to form a hybrid, and the simplest way to do that is to take a weighted average of the two:

(27) $\lambda \frac{\sum_{i=1}^{n} x_i}{n} + (1-\lambda) \prod_{i=1}^{n} x_i^{\frac{1}{n}} = k$, for some $\lambda \in [0,1]$

The problem with this scheme, however, is that the weight attached to each function is independent of the reserves $(x_1, x_2, \ldots . x_n)$. Ideally, we would like to assign more weight to the CMMM when any of these token reserve approaches zero to ensure its price escalates rapidly. Similarly, when the amounts of the tokens are roughly equal, assigning greater weight to the CSMM will ensure that prices do not change too rapidly, thereby minimizing slippage.

To construct a scheme where λ is sensitive to the quantity of reserves, we note the arithmetic mean (AM) of the reserves is $A = \frac{\sum_{i=1}^{n} x_i}{n}$. The CSMM $\frac{\sum_{i=1}^{n} x_i}{n} = k$, consequently,

[56] Specifically, Egorov (2019) uses the exchange function $\chi D^{n-1} \sum x_i + \prod x_i = \chi D^n + (\frac{D}{n})^n$, where χ is a 'leverage' parameter, D is the sum all token balances, x_i is the reserve of token i, and n is the number of tokens.

specifies various combinations of reserves that yield a particular AM. Similarly, the geometric mean (GM) of the reserves is $G = \prod_{i=1}^{n} x_i^{\frac{1}{n}}$; the CMMM $\prod_{i=1}^{n} x_i^{\frac{1}{n}} = k$ looks at combinations of reserves yielding a specific GM. Now, consider the relationship between A and G. A well-known theorem (which we do not prove here) is the AM-GM inequality:

The AM-GM inequality: For any non-negative real numbers $x_1, x_2, \ldots, x_n$:

(i) $\frac{\sum_{i=1}^{n} x_i}{n} \geq \prod_{i=1}^{n} x_i^{\frac{1}{n}}$

(ii) $\frac{\sum_{i=1}^{n} x_i}{n} = \prod_{i=1}^{n} x_i^{\frac{1}{n}}$ if and only if $x_1 = x_2 = \cdots = x_n$

Suppose we focus on the ratio G/A, where $A \neq 0$. Given the AM-GM inequality, it follows that $\frac{G}{A} \leq 1$ and that $\frac{G}{A} = 1$ only when the reserves of all tokens are equal. Moreover, it is evident that $\frac{G}{A} \geq 0$ and that $\frac{G}{A} = 0$ only if at least one of the reserves (but not all, since $A \neq 0$) satisfies $x_i = 0$. So, $0 \leq \frac{G}{A} \leq 1$ holds, and $\frac{G}{A}$ is a feasible candidate for the weight λ. Let us set, then, $\lambda = \frac{G}{A}$ in equation (27), which yields:

(28) $G\left(2 - \frac{G}{A}\right) = k$

From (28), we can derive the bilateral price using the same procedure as Section 4.1:

(29) $P_{X_j/X_i} = -\frac{dx_j}{dx_i} = \frac{2\left(1-\frac{G}{A}\right)\frac{\partial G}{\partial x_i} + \frac{G^2}{A^2}\frac{\partial A}{\partial x_i}}{2\left(1-\frac{G}{A}\right)\frac{\partial G}{\partial x_j} + \frac{G^2}{A^2}\frac{\partial A}{\partial x_j}}$

Equation (29) implies that when reserves of any token, say X_i, tends to zero, we have that $\frac{G}{A} \to 0$ and $P_{X_j/X_i} \to \frac{\partial G/\partial x_i}{\partial G/\partial x_j}$, which is the price that would exist under a CMMM. Similarly, when $\frac{G}{A} \to 1$, $P_{X_j/X_i} \to \frac{\partial A/\partial x_i}{\partial A/\partial x_j}$, which is the price that exists under a CSMM.

The outcome of this discussion is shown geometrically in Figure 8 for the two token case. The curve GG represents the CMMM $x^{\frac{1}{2}}y^{\frac{1}{2}} = k$ and the line AA represents the CSMM $\frac{x+y}{2} = k$. The mixing of the two AMMs using a fixed λ is shown with the dashed-curve CC, which has been drawn with $\lambda = 0.5$. It is evident that CC reduces the problem of dwindling reserves experienced by the CSMM AA; however, there is still considerable price slippage as, starting from the $45°$ line, the slopes change relatively rapidly. Choosing a different λ would involve a different trade-off between the twin problems of draining reserves and price slippage; the point is that this trade-off is fixed for all reserve combinations once λ is chosen. The HFMM with $\lambda = \frac{G}{A}$ is shown using the dotted-curve HH. Compared to the CC curve, HH is closer to AA when reserves are roughly equal and, at the same time, is closer to GG when one of the reserves in the liquidity pool starts to fall, thereby transferring weight to the AMM that better

resolves the more damaging problem for a specific reserve combination.

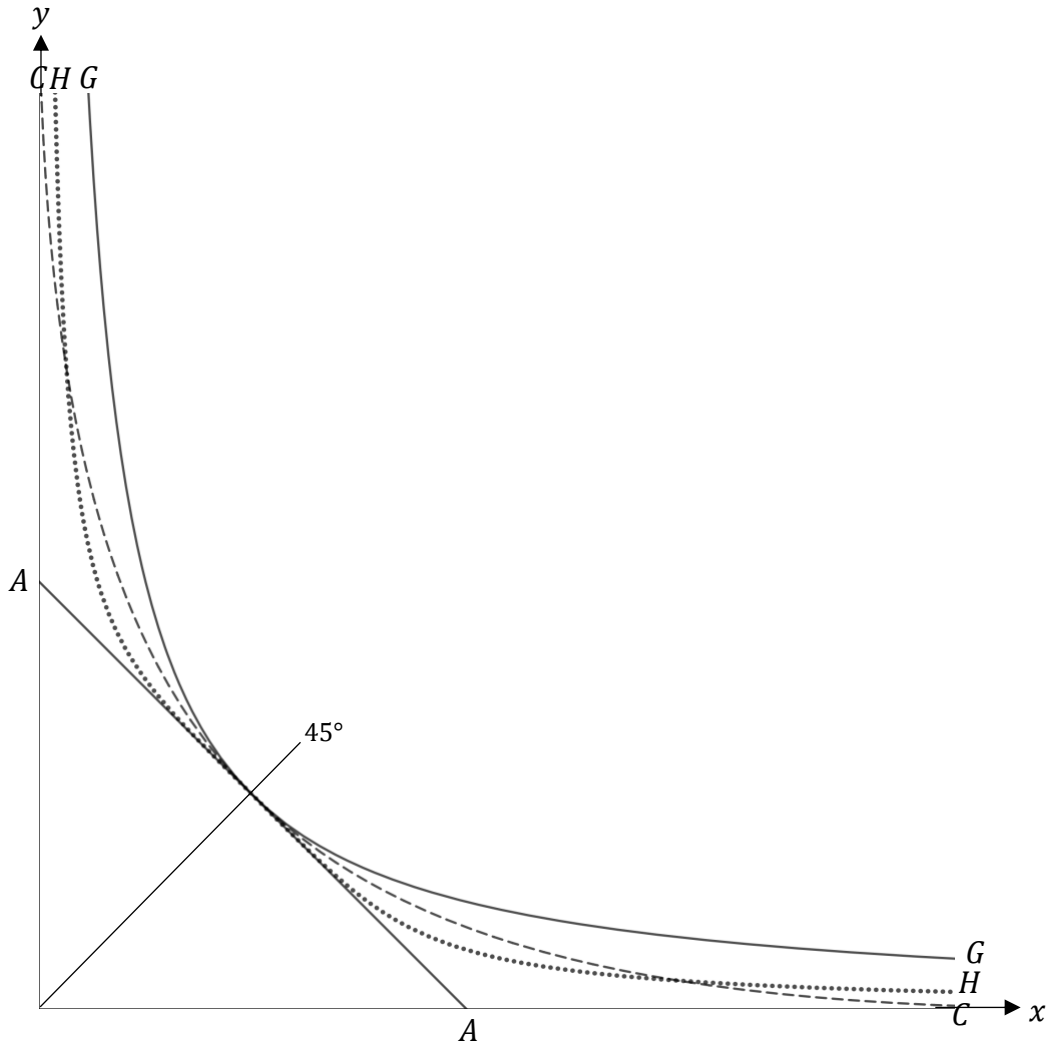

Figure 8: Hybrid Function Market Makers

While this example of forming a hybrid represents one way of moving between a CSMM and a CMMM, other possibilities exist. For example, one could specify the general CES (Constant Elasticity of Substitution) function as the AMM exchange function. The CES function with two tokens X and Y is $E[\alpha x^{\rho} + (1-\alpha)y^{\rho}]^{\frac{1}{\rho}}$, where E is an efficiency parameter, α is the distribution parameter and ρ is the substitution parameter. When $E = 1$, this will approach the two token CMMM exchange function $x^{\alpha}y^{1-\alpha}$ as $\rho \to 0$ and equals the CSMM $\alpha x + (1-\alpha)y$ when $\rho = 1$.[57] This allows for a way to change the curvature of the exchange function parametrically, similar to changing a fixed λ above.[58]

[57] See Chiang (1984) or Silberberg (1990) for a proof of convergence to $x^{\alpha}y^{1-\alpha}$ as $\rho \to 0$.

[58] When $E = 1$, $\alpha = 1/2$ and $\rho = 1/2$, the function becomes $(\frac{1}{2}x^{\frac{1}{2}} + \frac{1}{2}y^{\frac{1}{2}})^2$. When expanded this is $\frac{1}{2}A + \frac{1}{2}G$, the weighted average of A and G, with $\lambda = 0.5$. This yields the curve CC in Figure 8.

4.4 Dynamic Automated Market Makers (DAMM)

4.4.1 Dynamic weights in a CMMM

Section 4.3 provided the basic intuition behind how a HFMM attempts to modify a CMMM to keep prices relatively stable. A HFMM provides a satisfactory solution in the context of stablecoins where prices exhibit low volatility in the external market. However, the HFMM method of changing the shape of the AMM function may be less suitable when dealing with volatile tokens, where arbitrage activity may be high. Fees reduce arbitrage but negatively affect the efficiency of an AMM. So, the question that we now address is whether arbitrage activity on a CMMM be reduced: (a) while still reflecting the external market price; but, (b) without changing the $\prod_{i=1}^{n} x_i^{w_i} = k$ exchange function; and (c) without increasing fees?

In Section 4.1, the CMMM required assumed that the weights were fixed at some level. Consider, now, a scenario where the CMMM could alter the weights dynamically in response to some external market stimulus. To see the implications of this possibility, let us recap the CMMM arbitrage response with zero trading fees. If $M_{X_j/X_i} < P_{X_j/X_i} = \frac{w_i}{w_j}\frac{x_j}{x_i}$, for example, arbitrageurs purchase X_i in the external market and sell it in the CMMM in exchange for X_j till $M_{X_j/X_i} = P_{X_j/X_i}$ holds. All the response in this case is through changes in $\frac{x_j}{x_i}$. There is, however, an alternative way to get P_{X_j/X_i} to fall: by reducing $\frac{w_i}{w_j}$. If the CMMM were to adjust weights to w_i' and w_j', we can have $M_{X_j/X_i} = P_{X_j/X_i} = \frac{w_i'}{w_j'}\frac{x_j}{x_i}$, with no change in the ratio of reserves $\frac{x_j}{x_i}$. Prices in the two markets have become aligned not because of arbitrage activity, but rather due to 'dynamic' adjustments of the weights.

The above discussion suggests that weights that are flexible provides an additional equilibrating factor other than arbitrage to ensure that AMM prices track external market prices. The external market price now serves the purpose of an *oracle*: an external source of information that is fed into the smart contract. The relative importance of dynamic weights and arbitrage in keeping prices aligned between the two markets then depends on how frequently and reliably the oracle provides information to the AMM. Finally, it is worth noting that dynamic weights cause two important changes to the CMMM. First, the shares in equation (25) are no longer fixed, causing the CMMM to no longer be analogous to a standard fixed weight index fund. Second, from (24), in the standard CMMM, the valuation changes only when the quantity x_1 changes. Changing weights, however, present an additional source of change for pool valuation, potentially making the pool value more volatile.

4.4.2 An example of a DAMM: Bancor

Bancor is an example of an AMM that dynamically adjusts weights as the external market price changes, thereby reducing arbitrage activity. Bancor does not directly use a CMMM

function as its trading technology; our goal here is to show that the price output is, nevertheless, the same as the CMMM discussed above. The important institutional features of Bancor described here are based on Hertzog et al (2018) and Rosenfeld (2017).[59]

A smart contract in Bancor has two types of tokens: a 'smart token' (say Y) and a set of n 'connected tokens', where $n \geq 1$. Let the connected tokens be $X_1, X_2 \ldots, X_n$. Let P_{Y/X_i} denote the price of 1 unit of X_i in terms of Y (in the absence of any trading fees).[60] A trader can purchase or sell X_i against Y at any time using the smart contract at the going price. When a trader sells X_i its reserve balance, x_i, increases and the smart contract automatically increases supply of Y. The opposite occurs when a trader buys X_i: reserve x_i decreases, as does the supply of Y. The mechanics of operation thus far appear entirely different than a CMMM. What makes the two comparable are 'connector weights' or 'fractional reserve ratios', $w_i, i \in \{1,2, \ldots n\}$. Letting y denote the total quantity of Y, w_i satisfies the condition:

(30) $\quad w_i = \frac{P_{Y/X_i} x_i}{y}$ for all , $i \in \{1,2, \ldots n\}$

The numerator is the value of token X_i denominated in units of Y, while the denominator is the total value of Y in the smart contract. Equation (30) essentially defines the connector weight as the ratio of the value of reserves of X_i to the total value of all the Y supplied. When $\sum_{i=1}^{n} w_i = 1$, it follows that $\sum_{i=1}^{n} P_{Y/X_i} x_i = y$ and the sum of the values of all the 'connector tokens' exactly equals the value of Y minted by the smart contract.

Now, equation (30) can be rearranged to get P_{Y/X_i}:

(31) $\quad P_{Y/X_i} = \frac{y w_i}{x_i}$

For any two tokens, X_i and X_j, we can get the cross-rate that is implied by (31):

(32) $\quad P_{X_j/X_i} = \frac{P_{Y/X_i}}{P_{Y/X_j}} = \frac{w_i}{w_j} \frac{x_j}{x_i}$

This is, of course, exactly the same as equation (23) for a CMMM. So, any set of reserves on Bancor generates exactly the same price as a CMMM when the connector weight for each token on Bancor equals the weight for that token on the CMMM. Consequently, Bancor and Balancer can be viewed as being equivalent in terms of the technology for converting quantities of tokens reserves to prices, even though the institutional mechanisms may be different. It also follows straight away that when Bancor assigns dynamic weights, its outcomes replicate the CMMM with dynamic weights outlined in Section 4.4.1 and the intuition of price adjustment provided there is applicable here as well.

[59] Our exposition here is meant to capture the overall logic of the system, rather than all the institutional details of the Bancor protocol; see https://app.bancor.network/eth/data (accessed on 20th September, 2021), Hertzog et al (2018) and Rosenfeld (2017) for further details.

[60] Hertzog et al (2018) and Rosenfeld (2017) work with the price $P_{X_i/Y}$, which (from equation 2) is the reciprocal of P_{Y/X_i} used here; the latter is used in this paper mainly for expositional ease.

Overall, revisiting Figure 1, we have formed connections between a number of different types of AMMs: CPMM-v1 and v2, CMMM, HFMM, CSMM and DAMM. While the possibilities for experimenting with AMMs appear endless, simple geometric tools based on homogeneity and homotheticity are often suffice to discern similarities and differences in their structures.

5. An AMM with 'concentrated liquidity': Uniswap-v3

5.1 Proposed changes in Uniswap-v3

The Unswap-v3 whitepaper (Adams et al, 2021) proposes three important changes to v1 and v2, which we summarize briefly here.

First, in the Uniswap-v1 and v2 protocols, tokens of liquidity providers are pooled together and can be traded anywhere along the exchange function. In contrast, Uniswap-v3 implements *concentrated liquidity*, where a liquidity provider can specify a price range, $\left[\underline{P}, \overline{P}\right]$, within which to add liquidity. The main purpose of this change is to improve capital efficiency. Consequently, liquidity provision no longer requires deposits of X and Y of equal value. Since a liquidity provider needs to cover only the range specified, the deposit ratios of can vary depending on the price range specified and the current price of tokens.

The implication of introducing concentrated liquidity is that each position is unique. Intuitively, in Uniswap-v1 and v2, liquidity provision was characterized by the amount Δx added to the pool (the condition that tokens are provided in equal value determines Δy); as a result, fungible liquidity tokens are capable of tracking liquidity contributions by providers. In Uniswap-v3, however, liquidity provision is characterized by Δx and by a price range $\left[\underline{P}, \overline{P}\right]$ specified by the provider, and the latter distinguishes the contribution of one provider from another who also adds Δx but specifies a different price range. As a result, the second major change is that v3 protocol uses non-fungible tokens to track liquidity.

The third change occurs to trading fees and how these fees are distributed to liquidity providers. In Uniswap-v1 and v2, all pools have a uniform trading fee of .3% and these were automatically added to the liquidity pool causing k to expand over time from fees alone. In v3, multiple pools can exist for the same token pair with varying fees, with .05%, .3% or 1% being the initial fee tiers (with the capability to add more tiers). Moreover, fees are no longer added to the liquidity pool, and are stored separately. In some ways, the addition of fee tiers is an alternative strategy to *ex ante* designing an 'optimal' fee for the AMM; liquidity providers can select their preferred fees for any pair of tokens and the market can converge towards a trading fee for the token pair. One would expect that, over time, this will also be conditioned by the extent of competition that exists between AMMs providing similar services.

5.2 The geometry of a single concentrated liquidity position

We now investigate more closely the implications of introducing concentrated liquidity. As will be seen, the geometry becomes more complex due to this feature. Our analysis here

follows broadly along the lines of Adams et al (2021) and Mellow Protocol (2021). However, we develop a set of geometric techniques that are consistent with the previous sections, and that allows us to better understand the relationship between Unsiwap-v1 and v2 on one hand, and Unsiwap-v3 on the other. Trading fees are assumed to be zero for simplicity.

Consider the exchange function $k = xy$ in Figure 9 below, which is the same as the Uniswap-v2 exchange function. To begin with, we assume that there exists a single user who has specified the price interval $[P^F_{Y/X}, P^C_{Y/X}]$ for the liquidity bounds. As before, we have that $P^C_{Y/X} = \frac{y^C}{x^C}$ and $P^F_{Y/X} = \frac{y^F}{x^F}$. Assume the current price is $P^B_{Y/X} = \frac{y^B}{x^B}$. A position is characterized by $P^F_{Y/X}$, $P^C_{Y/X}$ and the reserves placed within that price range.[61]

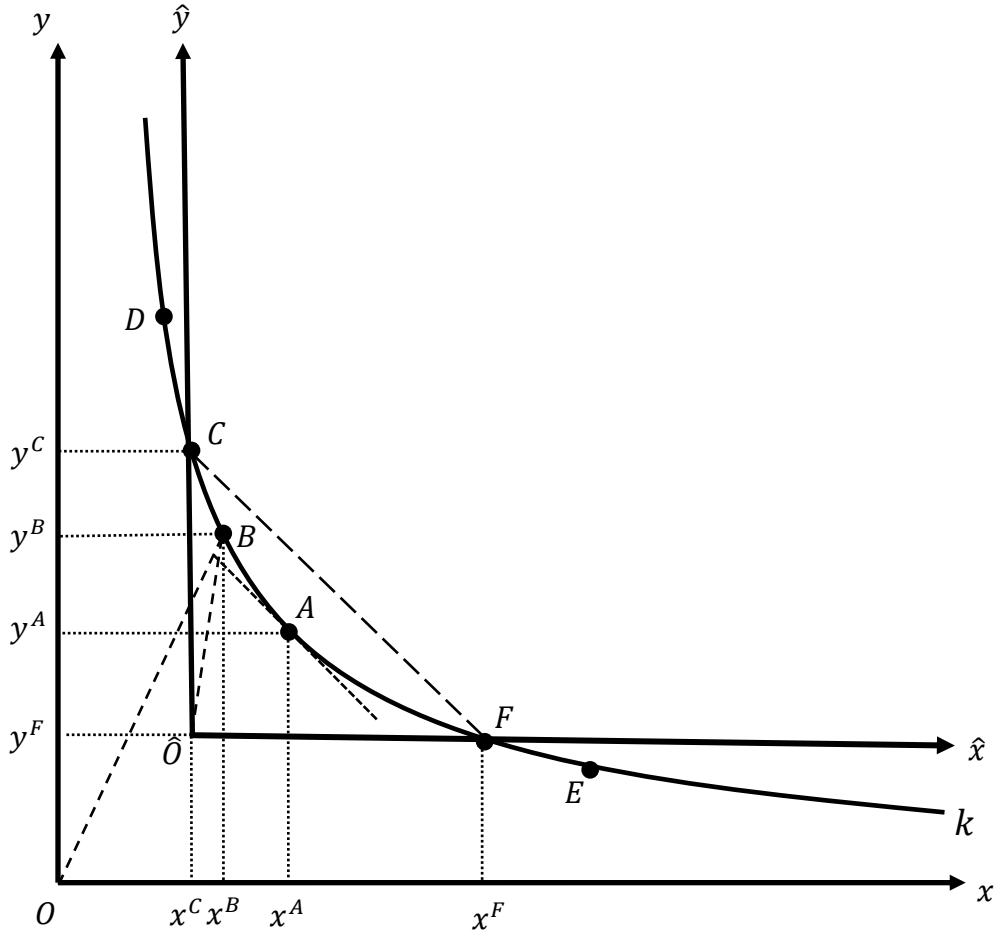

Figure 9: A single position in a Uniswap-v3 pool

There are now two aspects to focus on: first, what occurs in a position; and second, how this relates to the overall exchange function. To distinguish between the two, Uniswap-v3

[61] Uniswap-v3 keeps track of every .01% change between prices. Specifically, a pool tracks prices $P = 1.0001^i$, for $i \in \{\dots, -2, -1, 0, 1, 2, \dots\}$; each i corresponds to a price 'tick'. The change in prices between two ticks is then $\frac{dP}{di} = P\ln(1.0001) \cong .0001 \times P$. This results in a discretization of the prices, a complication that we shall conveniently ignore here.

uses the term *real* to keep track of actual reserves within a specific price range and *virtual* to refer to the overall exchange function. The exchange function $k = xy$ is defined, then, in terms of virtual reserves x and y. Virtual reserves are important in this context because they determine prices along the exchange function.

Geometrically, Figure 9 differentiates between virtual and real by drawing a separate set of 'real axes', $\hat{x}$ and $\hat{y}$ (with origin $\hat{O}$), to keep track of real reserves within the position CF. The 'virtual axes', x and y, keep track of virtual reserves. The real axes are, essentially, a translation of the virtual axes. Any point (x, y) in the virtual $x - y$ space can be translated to a point $(\hat{x}, \hat{y}) = (x - h, y - m)$ in the real $\hat{x} - \hat{y}$ space, where (h, m) is the co-ordinate of $\hat{O}$ in relation to the origin O. In Figure 9, $h = x^C$ and $m = y^F$; the coordinate of the real origin $\hat{O}$ is clearly a function of the bounds of the interval $[P^F_{Y/X}, P^C_{Y/X}]$ specified by the liquidity provider. Thus, at point C, the virtual reserves are x^C and y^C, but the position has real reserves $\hat{x}^C = (x^C - x^C) = 0$ of X and $\hat{y}^C = (y^C - y^F)$ of Y. Similarly, at point F, the real reserves are $\hat{x}^F = (x^F - x^C)$ and $\hat{y}^F = 0$ (while virtual reserves are y^F and x^F). In general, for any virtual reserve (x, y) in the arc FC, the translation of the axes yields corresponding real reserves $(\hat{x}, \hat{y}) = (x - x^C, y - y^F)$.

Now, suppose we start at some point, say B, where real reserves are equal to $(\hat{x}^B, \hat{y}^B) = (x^B - x^C, y^B - y^F)$. As traders swap X for Y, the liquidity pool loses Y and gains X, with prices moving along the exchange function till we hit point F. Here, $\hat{y}^F = 0$ and the position holds only X. Similarly, as prices move in the reverse direction from B, the position will run out of reserves of X at point C and be left with reserves of only token Y.

The real axes $\hat{x}$ and $\hat{y}$ highlight that even though the virtual reserves can never fall to zero for any asset due to the curvature of the $xy = k$ exchange function, real reserves can indeed do so. The latter occurs when the level curve intersects the $\hat{y}$ (or $\hat{x}$) axis. This necessitates differentiating between *active* and *inactive* positions: when the actual trading price is in the range $(P^F_{Y/X}, P^C_{Y/X})$, the position CF is active. This occurs, for example, when the actual price equals the slope of ray OB in Figure 9. When the price is outside that range, at say points D or E, the position is inactive. Inactive positions earn no fees on Uniswap-v3, which is a marked difference from Uniswap-v2 where liquidity providers earn fees along the entire exchange function. If the position is inactive at the time of liquidity provision, say at point D, the liquidity provider deposits only asset Y when specifying the inactive range $[P^F_{Y/X}, P^C_{Y/X}]$. Similarly, if the current price corresponds to point E, the liquidity provider deposits only asset X when specifying the same price range. In either case, the position becomes active when the price enters the specified price range.

The feature of real reserves falling to zero bears similarity to a CSMM, such as the one shown in Figure 7. To highlight the similarity, we have drawn the chord connecting C and F in Figure 9; with respect to the axes $\hat{x}$ and $\hat{y}$, the line CF behaves *like* a CSMM. As pointed

out in Section 4.2, the slope of the CSMM is the fixed price at which the two assets trade. It is worth investigating, then, what information the slope of the line CF provides in Uniswap-v3. The slope of line CF is $\frac{y^C-y^F}{x^C-x^F}$, which is the rate at which the two assets exchange *on average* in the position. We can, therefore, think of the absolute value of the slope of CF as the average price of the liquidity position bounded by the prices $P^F_{Y/X}$ and $P^C_{Y/X}$. Denoting the average price as $P^{CF}_{Y/X}$:

(33) $$P^{CF}_{Y/X} = \frac{y^C-y^F}{x^F-x^C}$$

To understand how the chord CF is related to the arc CF, we utilize a fundamental theorem in calculus, the Mean Value Theorem.

Mean Value Theorem: Consider a continuous function $y = f(x)$ on an interval $[x_1, x_2]$ that is differentiable in (x_1, x_2). Then, there exists some $x_A \in (x_1, x_2)$ such that:

$$\frac{df(x_A)}{dx} = \frac{f(x_2)-f(x_1)}{x_2-x_1}$$

In terms of Figure 9, the Mean Value Theorem implies that there exists a point A (as shown in the figure) on the exchange function between C and F that has the same slope as the line CF. Consequently, we can identify A as a point on the exchange function that has the same price as the average price of the liquidity position bounded by C and F. It is straightforward to show that this average price is the geometric mean of the price bounds:[62]

(34) $$P^{CF}_{Y/X} = \sqrt{P^F_{Y/X} P^C_{Y/X}}$$

The average price is also the effective price for a *range order* in Unsiwap-v3, where a liquidity provider contributes a single token in an inactive position (typically with a narrow price range). In Figure 9, if the liquidity position becomes active at F and the price moves through arc CF before exiting at C, the average price for exchanging X to Y is then given by (34).

It is worth noting that while "price equals the ratio of reserves" is an unambiguous statement in Uniswap v1 and v2, this is no longer true in Uniswap-v3, where there are two reserve ratios to keep track of. First, the virtual reserve ratio $\frac{y}{x}$ (which equals $P_{Y/X}$) and second, the real reserve ratio $\frac{\hat{y}}{\hat{x}}$ (which need not). To see the difference between the two, in Figure 9 at point B, we have that the virtual reserve ratio $\frac{y^B}{x^B}$ equals the slope of the ray OB. However, the real reserve ratio is $\frac{\hat{y}^B}{\hat{x}^B} = \frac{y^B-y^F}{x^B-x^C}$, which is obviously different. Indeed, the real

[62] Specifically, we have that $P^{CF}_{Y/X} = \frac{y^C-y^F}{x^F-x^C} = \frac{k(\frac{1}{x^C}-\frac{1}{x^F})}{x^F-x^C} = \frac{k}{x^F x^C} = \sqrt{\frac{k^2}{(x^F x^C)^2}} = \sqrt{\frac{y^F y^C}{x^F x^C}} = \sqrt{P^F_{Y/X} P^C_{Y/X}}$.

reserve ratio equals the slope of the line $\hat{O}B$ running through the real origin. The real reserve ratio $\frac{\hat{y}}{\hat{x}}$ ranges from 0 to ∞ as the virtual reserve ratio $\frac{y}{x}$ varies between $P^F_{Y/X}$ and $P^C_{Y/X}$.

The upshot of the above discussion is that all these variables lying on the same graph - $x, y, \hat{x}, \hat{y}, P_{Y/X}, P^F_{Y/X}, P^C_{Y/X}, k$ – must be systematically linked. To delve into these links, we follow Adams et al (2021) by substituting $L = \sqrt{k}$ in the exchange function, though this is not necessary and one can simply use $\sqrt{k}$ instead of L in subsequent analysis.[63] With this change, the exchange function is $xy = L^2$. For a given interval $[P^F_{Y/X}, P^C_{Y/X}]$, the virtual and real reserves are linked by $x = \hat{x} + x^C$ and $y = \hat{y} + y^F$. The exchange function is then:

(35) $$(\hat{x} + x^C)(\hat{y} + y^F) = L^2$$

Once the liquidity provider has specified interval $[P^F_{Y/X}, P^C_{Y/X}]$, the quantities x^C and y^F are fixed along a level curve. This allows us to describe x^C and y^F in terms of $L, P^F_{Y/X}$ and $P^C_{Y/X}$. Specifically, we get that $x^C = \frac{L}{\sqrt{P^C_{Y/X}}}$ and $y^F = L\sqrt{P^F_{Y/X}}$.[64] Plugging this in equation (35) yields:

(36) $$(\hat{x} + \frac{L}{\sqrt{P^C_{Y/X}}})(\hat{y} + L\sqrt{P^F_{Y/X}}) = L^2$$

Equation (36) is useful because it tells us the various combinations of real reserves $(\hat{x}, \hat{y})$ that yield the same level of L for given $P^C_{Y/X}$ and $P^F_{Y/X}$. In other words, it describes the behavior of real reserves in the arc FC in Figure 9.

Next, consider any arbitrary point on the arc FC characterized by virtual reserves (x, y) and price $P_{Y/X} = \frac{y}{x}$. It is worthwhile to describe real reserves at that point, $(\hat{x}, \hat{y}) = (x - x^C, y - y^F)$, in terms of $L, P^F_{Y/X}, P^C_{Y/X}$ and $P_{Y/X}$:[65]

(37) $$\hat{x} = \frac{L}{\sqrt{P_{Y/X}}} - \frac{L}{\sqrt{P^C_{Y/X}}} \quad \text{and} \quad \hat{y} = L\sqrt{P_{Y/X}} - L\sqrt{P^F_{Y/X}}$$

Equation (37) allows us to see how logically consistent liquidity provision occurs on Uniswap-v3 for an active position, when the current price is $P_{Y/X}$. Eliminating L in (37), we get:

(38) $$\hat{y} = \hat{x}\left(\frac{\sqrt{P_{Y/X}} - \sqrt{P^F_{Y/X}}}{\frac{1}{\sqrt{P_{Y/X}}} - \frac{1}{\sqrt{P^C_{Y/X}}}}\right)$$

[63] However, as pointed out in Section 3.4.2, $L = \sqrt{k}$ is a better measure of liquidity. The idea of adding liquidity from different positions is more intuitively captured using L.

[64] At any arbitrary point (x, y), we have that $xy = L^2$ and $\frac{y}{x} = P_{Y/X}$, which implies that $x = \frac{L}{\sqrt{P_{Y/X}}}$ and $y = L\sqrt{P_{Y/X}}$. Since this is true at any arbitrary point, it holds at both points C and F.

[65] This follows directly from the previous footnote.

So, if a user specifies a certain quantity $\hat{x}$ of token X, along with a price range bounded by $P^C_{Y/X}$ and $P^F_{Y/X}$, equation (38) determines the amount of token Y that must be added by the user for liquidity provision at the current price of $P_{Y/X}$. To see how this compares to Uniswap-v1 and v2, recall that in these earlier versions, liquidity is provided over the entire range of the exchange function. This is equivalent to a position on Uniswap-v3 where $P^F_{Y/X} \longrightarrow 0$ and $P^C_{Y/X} \longrightarrow \infty$; indeed, when these limits are taken in equation (38), we get $\hat{y} = \hat{x}P_{Y/X}$, which is the Uniswap-v1 and v2 liquidity provision rule.

5.3 Combining positions

Section 5.2 showed that for liquidity provision in a single position, given the current price ($P_{Y/X}$) and a user-defined price range (bounded by $P^C_{Y/X}$ and $P^F_{Y/X}$), a specified amount $\hat{x}$ of token X is associated with an amount $\hat{y}$ of token Y (equation (38)) and a certain level curve $k = L^2$ (equation (37)). This can be now used to derive how multiple positions interact for liquidity provision on Uniswap-v3.

Consider, to begin with, the simplest possible case when there are two positions defined over the same price interval $[P^F_{Y/X}, P^C_{Y/X}]$, but where the liquidity providers specify two separate quantities of token X when providing liquidity: $\hat{x}'$ and $\hat{x}''$. At the current price of $P_{Y/X}$, the amounts of token Y that must be added is given by equation (38). From equation (37), we see that these amounts of X are consistent with different level curves:

(39) $$L' = \frac{\hat{x}'}{\frac{1}{\sqrt{P_{Y/X}}} - \frac{1}{\sqrt{P^C_{Y/X}}}} \quad \text{and} \quad L'' = \frac{\hat{x}''}{\frac{1}{\sqrt{P_{Y/X}}} - \frac{1}{\sqrt{P^C_{Y/X}}}}$$

It is readily observed that summing these positions is equivalent to a single position with $\hat{x}''' = \hat{x}' + \hat{x}''$; that is:

(40) $$L' + L'' = \frac{\hat{x}'}{\frac{1}{\sqrt{P_{Y/X}}} - \frac{1}{\sqrt{P^C_{Y/X}}}} + \frac{\hat{x}''}{\frac{1}{\sqrt{P_{Y/X}}} - \frac{1}{\sqrt{P^C_{Y/X}}}} = \frac{\hat{x}'''}{\frac{1}{\sqrt{P_{Y/X}}} - \frac{1}{\sqrt{P^C_{Y/X}}}} = L'''$$

This is shown graphically in Figure 10 below, where the real axes for the two positions have been drawn with origins $\hat{O}'$ and $\hat{O}''$. The prices $P^F_{Y/X}$, $P^C_{Y/X}$ and $P_{Y/X}$ correspond to the slopes of rays OF, OC and OP, respectively. The amount of X provided in the first position, $\hat{x}'$, corresponds to the distance $\hat{O}'\hat{x}'$. Once the user has determined the amount of X to provide, the corresponding amount of Y is given by equation (38) and is shown by the distance $\hat{O}'\hat{y}'$. The level curve corresponding to these real reserves can be determined by equation (39) and is shown as $k' = (L')^2$. A similar set of arguments hold for the second position labelled with '' on the graph. The combined liquidity provision is given by equation (40), and is depicted by the level curve $k''' = (L''')^2$, where $L''' = L' + L''$. The amount of X in the combined position is given by $\hat{x}''' = \hat{x}' + \hat{x}''$ (the distance $\hat{O}'''\hat{x}'''$), and the amount of Y is $\hat{y}''' = \hat{y}' + \hat{y}''$ (the

distance $\hat{O}'''\hat{y}'''$). Once the liquidity is provided, as the current price changes from $P_{Y/X}$ when traders swap tokens, we move from A along k''' till point B or D is reached, beyond which the combined position becomes inactive and contains real reserves of only one token: X (at point D) or Y (at point B).

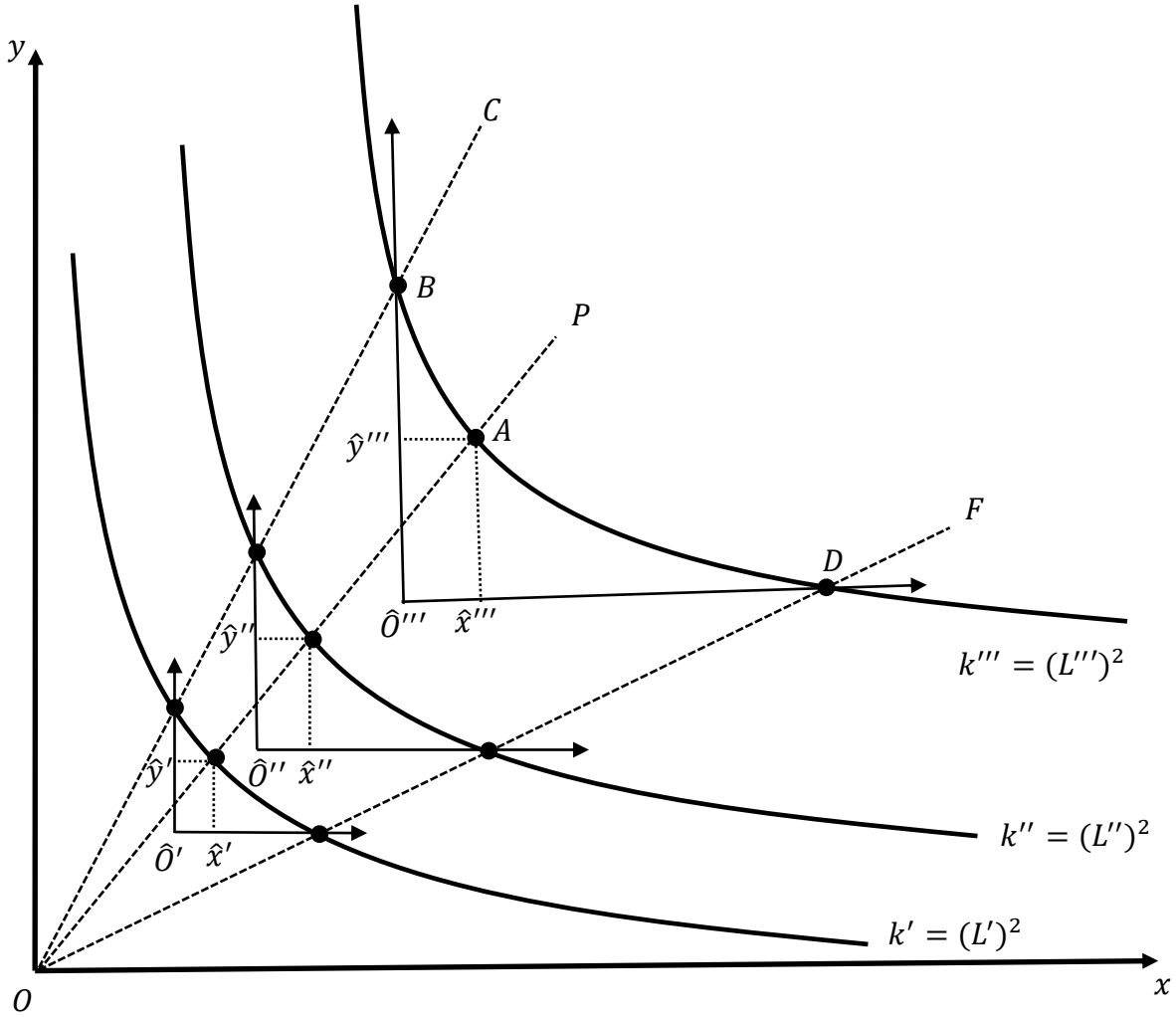

Figure 10: Combining positions on Uniswap-v3

In general, intervals specified by liquidity providers may partially overlap (or not at all). While this yields more complex graphs, the underlying logic is the same. In Figure 11 below, we depict a scenario with a partial overlap, with the two user-specified price ranges being $[P_{Y/X}^{F'}, P_{Y/X}^{C'}]$ and $[P_{Y/X}^{F''}, P_{Y/X}^{C''}]$. These price ranges and the corresponding liquidities translate the axes to $\hat{O}'$ and $\hat{O}''$, respectively. The prices can be read off the graph as the slope of rays OF', OF'', OC' and OC''. To keep the graph devoid of clutter, the figure does not contain information such as the current price, the real reserves and so on.

The overlap of the price interval suggests that there are 3 disjoint ranges to consider: $[P_{Y/X}^{F'}, P_{Y/X}^{F''}]$, $[P_{Y/X}^{F''}, P_{Y/X}^{C'}]$ and $[P_{Y/X}^{C'}, P_{Y/X}^{C''}]$. While the algebra of the breakdown is somewhat messy, the intuition of what happens in this instance is clear from the geometry. Suppose we start at a situation where the current price satisfies $P_{Y/X} < P_{Y/X}^{F'}$; at such a price both positions are inactive and real reserves are held entirely in token X. When the price increases to $P_{Y/X} = P_{Y/X}^{F'}$, the first position is activated. This occurs at point A in Figure 11.

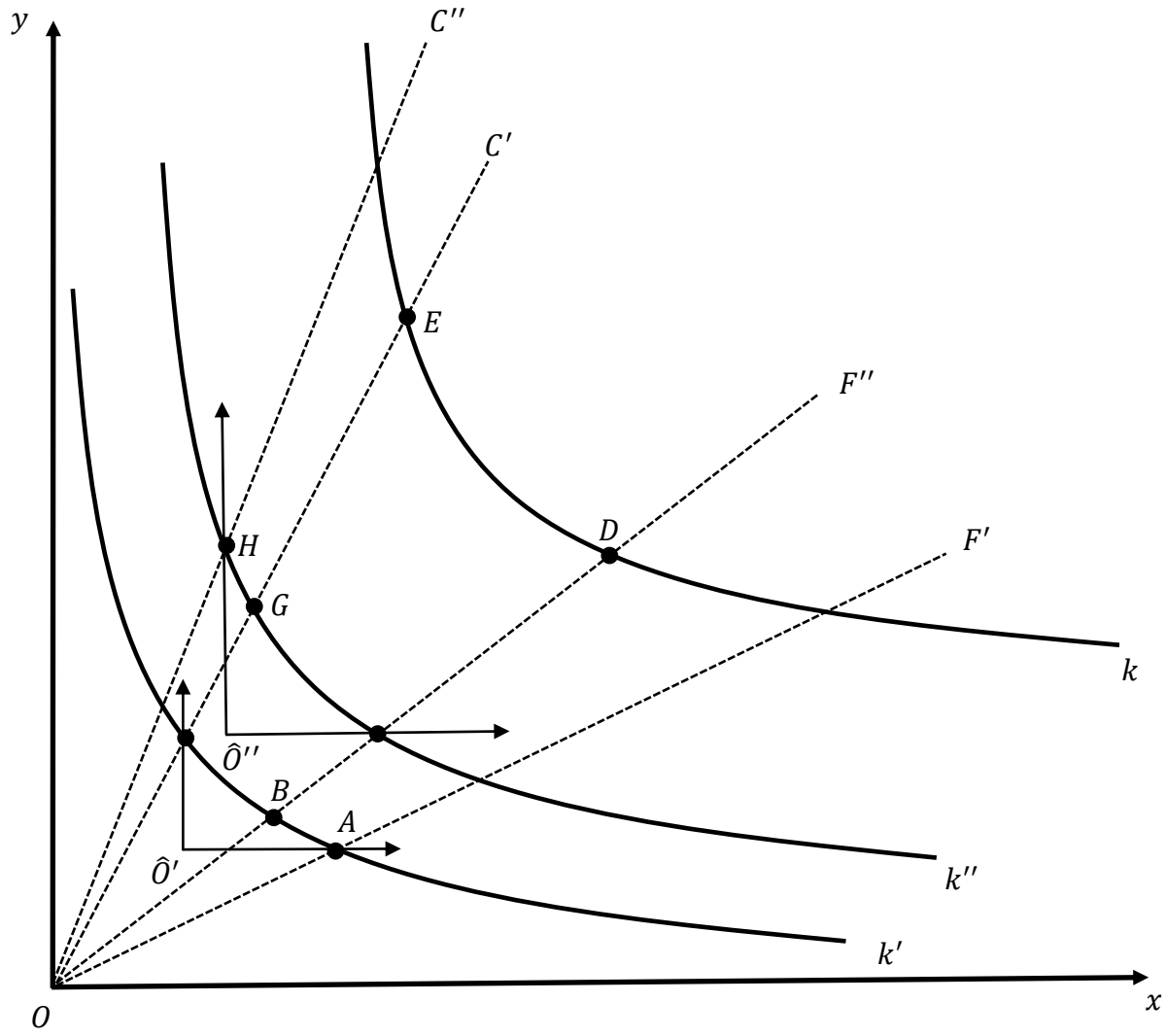

Figure 11: Overlapping positions on Uniswap-v3

In the price range $[P_{Y/X}^{F'}, P_{Y/X}^{F''}]$, only this position provides liquidity, so the relevant level curve is k'. This holds till point B, but once the price reaches $P_{Y/X} = P_{Y/X}^{F''}$ the second position is activated. Consequently, in the range $[P_{Y/X}^{F''}, P_{Y/X}^{C'}]$, both positions are active, and the relevant level curve of the exchange function is k, which combines both positions in that price range. There is, in essence, a jump from point B to D to reflect the joining of a second position

to liquidity provision. As the price increases beyond $P_{Y/X}^{F''}$, we move along the arc DE; at point E, we have that $P_{Y/X} = P_{Y/X}^{C'}$ and all the real reserves in the first position have been converted to token Y, leaving only the second position with liquidity in both tokens to facilitate further swaps by traders that involve the purchase of X. In the final range $[P_{Y/X}^{C'}, P_{Y/X}^{C''}]$, only the second position is active, so we drop to point G on the level curve k''. Further price increases involve moving along the arc GH, till $P_{Y/X} = P_{Y/X}^{C''}$, at which point real reserves of X are exhausted in the second position as well.

There are two features to take away from this discussion. First, unlike Uniswap-v1 and v2, trade does not take place along a single level curve. Rather, as shown by Figure 11, there are jumps in the level curves as trading occurs, leading to discontinuities. In Figure 11, there are three discontinuous arcs that characterize liquidity provision: arc AB, arc DE and arc GH. Secondly, despite this added complexity, the geometry behind Figure 11 is straightforward, and combining additional positions is readily visualized using the same procedure.[66]

Unsiwap-v3, therefore, functions differently than its predecessor v2 in significant ways. While the same exchange function is used to convert (now virtual) reserves to prices, the introduction of concentrated liquidity produces features that are distinct. Apart from the liquidity provisioning being discontinuous, each position behaves similar to a CSMM in the sense that it can run out of one reserve or the other, which cannot occur on Uniswap-v1 an v2. It is intuitive that adding concentrated liquidity to other AMMs, such as a CMMM like Balancer, will also produce such changes. As a consequence, it may be worthwhile to add to the alphabet soup of AMM labels by giving AMMs with concentrated liquidity their own name, perhaps Concentrated Liquidity Market Makers (CLMMs).[67]

Having pointed out how different Uniswap-v3 is to v2 and v1 in terms of yielding discontinuous liquidity levels, we conclude our analysis of Uniswap-v3 by summarizing when its liquidity provision does correspond to continuous curves in a manner similar to earlier versions. The first is when all liquidity providers specify a price range where $P_{Y/X}^{F} = 0$ and $P_{Y/X}^{C} = \infty$; this replicates Uniswap-v2 essentially because concentrated liquidity is no longer relevant. The second is when the level curve is partitioned by positions; partitioning implies that positions are disjoint and cover the entire range of prices. This is shown in Figure 12 below, where there are four disjoint positions, all of which involve liquidity provision such that they lie on the level curve k. Here, trading occurs along a single level curve with no discontinuities, like Uniswap-v2. However, each position can be drained of a single reserve, like a CSMM (dashed lines). Consequently, Figure 12 highlights the somewhat hybrid nature

[66] Though, as mentioned earlier, since price moves in ticks on Uniswap-v3 and fees are charged as well, the actual computations are more complicated than those described here.

[67] Cyclos, for example, labels itself as a concentrated liquidity market maker for the Solana network; see https://cyclos.io (accessed 20th September, 2021).

of Unsiwap-v3.

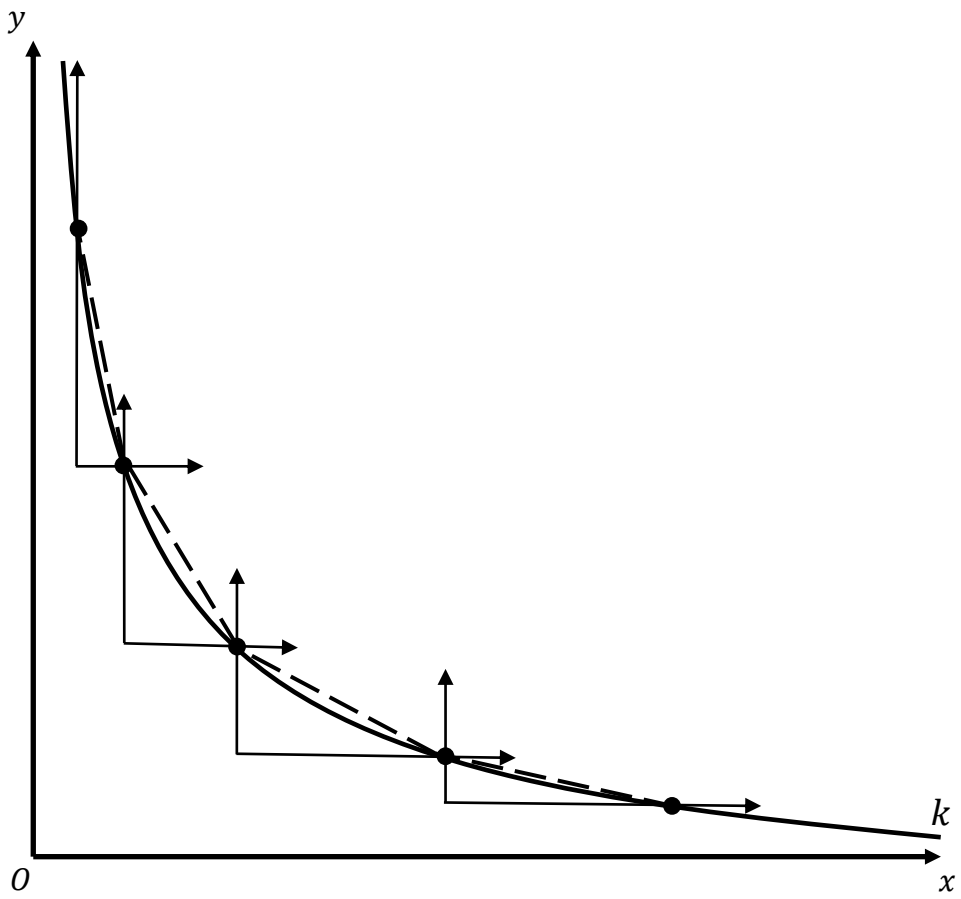

Figure 12: Partitioning the level curve on Unsiwap-v3

6. Conclusions

This paper presented a unified framework based on the neoclassical black-box to characterize different types of AMMs that are currently popular as DEXs. One of the main advantages of such a framework is that it provides a set of simple tools that can be used to visualize the geometry of a given AMM. This makes it easy to see, for example, why the price remains unchanged when both reserves are doubled. It also allows for a simple way to check when manipulating a given exchange function makes a significant difference and when it does not. In a CPMM, for example, the homotheticity property suggests that any monotonic transformation of the exchange function does not alter the price for any given set of reserves, making these various functions equivalent. Other properties, such as Euler's theorem provide a convenient method to value a liquidity pool in an AMM.

Importantly, the methodology also allows for comparisons across a variety of AMMs, in order to examine how similar or different various types of AMMs are. While the coverage of AMMs in this paper is certainly not exhaustive, especially since this is a fertile area for innovation with new ideas for AMMs are being developed at a rapid pace, the analytical and

geometric tools described here are fairly general and (hopefully) present an intuitive framework to allow new AMMs to be scrutinized and contrasted with existing AMMs.

The focus of this paper has been on the analytical framework behind AMMs. As such, AMMs rely on arbitrage or oracles to align prices with the broader market. Our understanding of how well arbitrage operates in this space and how it performs relative to oracles is still nascent. This is partly an empirical endeavour (which has not been addressed in this paper at all), but there are theoretical considerations here as well. For example, there are many impediments (other than transaction fees) to how efficiently AMMs can operate. These include phenomena such as frontrunning of trades, the worrisome presence of generalized bots that are capable of traversing the landscape of the Ethereum terrain to 'snipe' arbitrage transactions entered into by others, the reordering of transactions by miners, and so on.[68] The last couple are particularly interesting, because they can destroy incentives to engage in arbitrage: if arbitrage opportunities spotted by an agent (in the broadest of being any entity, human or a program, that recognizes an arbitrage opportunity) are likely to be stolen by a bot, there is little incentive to identify arbitrage opportunities in the first place. This questions the incentive structure for arbitrage to be undertaken.

While this primer takes a fairly deep look at AMMs and DEXs, it barely scratches the surface of advancements to financial instruments and institutions that are taking place in the DeFi space. This paper does not, for example, examine borrowing and lending institutions being developed, the use of derivatives such as perpetual swaps, or how stablecoins operate.[69] Moreover, as mentioned in the introduction, the main disadvantage of the black-box approach is that it precludes an examination of many interesting issues, such as how communities are formed in a decentralized environment, how governance takes place, what restrictions are imposed by the ability to fork, and so on. In other words, there is much scope for exciting research to be done in the DeFi space.

Like blockchains in general, DeFi is the amalgamation of different fields of expertise: computer science, cryptography, finance, economics and game theory, to name a few. These fields have developed with their own jargon, methodologies and tools of analysis, with ever growing levels of sub-specialization and narrowness in expertise.[70] It comes as no surprise,

[68] See Daian et al (2019) for these and other issues related to vulnerabilities in the DeFi space. Robinson and Konstantopoulos (2020) provide an interesting account of the Ethereum as a 'dark forest'.

[69] For a broader treatise on the current state of DeFi, see Harvey et al (2021).

[70] In a memorable exchange between Sherlock Holmes and Dr. Watson in *A Study in Scarlet*, when the latter expresses astonishment at the former's lack of awareness of the solar system, Holmes responds, "A fool takes in all the lumber of every sort that he comes across, so that the knowledge which might be useful to him gets crowded out, or at best is jumbled up with a lot of other things, so that he has a difficulty in laying his hands upon it. Now the skilful workman is very careful indeed as to what he takes into his brain-attic…" When Dr. Watson persists, "But the solar system!", Holmes counters with "What the deuce is it to me?", and then goes on to state, "You say that we go round the sun. If we went round the moon it would not make a pennyworth of difference to me or to my work."

then, that computers scientists and entrepreneurs at the forefront of the technological developments in the blockchain space often invent new jargon and methodologies to describe phenomena that economists have examined over many decades. At best this creates confusion; at worst, it creates inefficiencies in the scientific process as one field reinvents the knowledge already accumulated by another, or as one field eschews incorporating useful new ideas due to the barriers imposed by multidisciplinary communication. This review presents an attempt to connect advancements in DeFi with the traditional toolkit of economics for one use-case of DeFi, and in doing so takes a few, admittedly small, steps forward in the direction of developing an integrated theory of DeFi.

References

Adams, H., Zinsmeister, N. and Robinson, D. 2020. Unsiwap v2 Core. https://uniswap.org/whitepaper.pdf (accessed 20th September, 2021).

Adams, H., Zinsmeister, N., Salem, M., Keefer, R. and Robinson, D. 2021. Uniswap v3 Core. https://uniswap.org/whitepaper-v3.pdf (accessed 20th September, 2021).

Angeris, G., Kao, H., Chiang, R., Noyes, C. and Chitra, T. 2019. An Analysis of Uniswap Markets. https://arxiv.org/abs/1911.03380 (accessed on 20th September, 2021).

Angeris, G. and Chitra, T. 2020. Improved Price Oracles: Constant Function Market Makers. In *Proceedings of the 2nd ACM Conference on Advances in Financial Technologies* (AFT '20). Association for Computing Machinery, New York, NY, USA, 80–91. DOI: https://doi.org/10.1145/3419614.3423251

Buterin, V. 2014. A Next Generation Smart Contract and Decentralized Application Platform. https://cryptorating.eu/whitepapers/Ethereum/Ethereum_white_paper.pdf (accessed 20th September, 2021).

Buterin, V. 2018. Improving Front Running Resistance in x*y=k Market Makers. https://ethresear.ch/t/improving-front-running-resistance-of-x-y-k-market-makers/1281 (accessed 20th September, 2021)

Chiang, A.C. 1984. *Fundamental Methods of Mathematical Economics.* McGraw-Hill Inc., USA.

Daian, P., Goldfeder, S., Kell, T., Li, Y., Zhao, X., Bentov, I., Breidenbach, L., Juels, A. 2019. Flash Boys 2.0: Frontrunning, Transaction Reordering and Consensus Instability in Decentralized Exchanges. https://arxiv.org/pdf/1904.05234.pdf (accessed 20th September, 2021).

Demsetz, H. 1997. The Firm in Economic Theory: A Quiet Revolution. *The American Economic Review*, 87: 426 – 429.

Dernburg, T.F. 1989. *Global Macroeconomics*. New York: Harper Collins Publishers, Inc.

Egorov, M. 2019. StableSwap – Efficient Mechanism for Stablecoin Liquidity. https://curve.fi/files/stableswap-paper.pdf (accessed 20th September, 2021).

Hanson, R. 2003. Combinatorial Information Market Design. *Information Systems Frontiers,* 5: 107 – 119.

Hanson, R. 2007. Logarithmic Market Scoring Rules for Modular Combinatorial Information Aggregation. *The Journal of Prediction Markets*, 1: 3 – 15.

Hart, O. 1995. *Firms, Contracts and Financial Structure*. Oxford University Press, New York.

Hertzog, E., Benartzi, G. and Benartzi, G. 2018. Bancor Protocol. https://storage.googleapis.com/website-bancor/2018/04/01ba8253-bancor_protocol_whitepaper_en.pdf (accessed 20th September, 2021).

Holmstrom, B. and Tirole, J. 1989. The Theory of the Firm. In R. Schmalensee and R.D. Willig (eds.), *Handbook of Industrial Organization*, vol 1, *Handbooks in Economics*, no. 10, Amsterdam: North Holland, 61 – 133.

Martinelli, F. and Mushegian, N. 2019. Balancer: A Non-custodial Portfolio Manager, Liquidity

Provider, and Price Sensor. https://balancer.finance/whitepaper/ (accessed 20th September, 2021).

Mas-Colell, A. Whinston, M.D. and Greem, J.R. 1995. *Microeconomic Theory*. New York: Oxford University Press.

Mellow Protocol. 2021. Uniswap V3: Liquidity Providing 101. https://mellowprotocol.medium.com/uniswap-v3-liquidity-providing-101-f1db3822f16d (20th September, 2021)

Moosa, I. 2010. *International Finance*: *An Analytical Approach* (3rd edition). Sydney: McGraw-Hill Australia Pty Ltd.

Pintail. 2019. Uniswap: A Good Deal for Liquidity Providers? https://pintail.medium.com/uniswap-a-good-deal-for-liquidity-providers-104c0b6816f2 (accessed 20th September, 2021)

Pourpouneh, M. Nielsen, K. and Ross, O. 2020. Automated Market Makers. IFRO Working Paper 2020/08, available at https://www.econstor.eu/handle/10419/222424 (accessed 20th September, 2021).

Robinson, D. and Konstantopoulos, G. 2020. Ethereum is a Dark Forest. https://www.paradigm.xyz/2020/08/ethereum-is-a-dark-forest/ (accessed 20th September, 2021).

Rosenfeld, M. 2017. Formulas for Bancor System. http://meissereconomics.com/assets/abfe-lesson5-bancor.pdf (accessed 20th September, 2021).

Schär, F. 2021. Decentralized Finance: On Blockchain and Smart Contract-based Financial Markets. *Federal Reserve Bank of St. Louis Review*, 103: 153 – 174. https://doi.org/10.20955/r.103.153-74.

Silberberg, E. 1990. *The Structure of Economics: A Mathematical Analysis.* McGraw Hill, Inc., USA.

Szabo, N. 1996. Smart Contracts: Building Blocks for Digital Markets. https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart_contracts_2.html (accessed 20th September, 2021).

Tassy, M. and White, D. 2020. Growth Rate of a Liquidity Provider's Wealth in $XY = c$ Automated Market Makers. https://math.dartmouth.edu/~mtassy/articles/AMM_returns.pdf (accessed 20th September, 2021).

Varian, H. 1987. *Intermediate Microeconomics*: *A Modern Approach*. New York: W.W. Norton.

White, D., Tassy, M., Noyes, C. and Robinson, D. 2020. Uniswap's Financial Alchemy. https://research.paradigm.xyz/uniswaps-alchemy (accessed 20th September, 2021).

Xu, J., Vavryk, N., Paruch, K. and Cousaert, S. 2021. SoK: Decentralized Exchanges (DEX) with Automated Market Maker (AMM) protocols. https://arxiv.org/abs/2103.12732 (accessed 20th September, 2021).

Zhang, Y., Chen, X. Park, D. 2018. Formal Specification of Constant Product ($xy = k$) Market Maker Model and Implementation. https://github.com/runtimeverification/verified-smart-contracts/blob/uniswap/uniswap/x-y-k.pdf (accessed 20th September, 2021).

Zhou, L., Qin, K., Ferreira Torres, C., Gervais, A. 2020. High-Frequency Trading on Decentralized On-Chain Exchanges. https://arxiv.org/abs/2009.14021 (accessed 20th September, 2021).